Primate researchers divided over retirement for lab monkeys p. 1182

New books for young scientists p. 1186

Enzyme cascades streamline drug production pp. 1199 & 1255

Science

$15
6 DECEMBER 2019
sciencemag.org
AAAS

VOLCANIC SURPRISE

Distant eruptions trigger caldera collapse

pp. 1200, 1212, 1213, & 1214

CONTENTS

6 DECEMBER 2019 • VOLUME 366 • ISSUE 6470

1182
Retired from research, a macaque lives out its days at the Peaceable Primate Sanctuary in Winamac, Indiana.

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

BOOKS *ET AL.*

PERSPECTIVES

POLICY FORUM

LETTERS

RESEARCH

ON THE COVER

The Halema'uma'u crater lava lake illuminates a plume within Kīlauea's volcanic summit. The 2018 volcanic eruption and caldera collapse event at the summit were monitored using satellite and ground-based methods. The observations were important for real-time hazard assessment and response during the several-month-long eruption. See pages 1200, 1212, 1213, and 1214.
Photo: Mario Tama/Getty Images

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2019 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1971; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: CHRIS KOLACZAN/SHUTTERSTOCK

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri SR. EDITORIAL FELLOW Andrew M. Sugden (UK) SR. EDITORS Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Paula A. Kiberstis, Marc S. Lavine (Canada), Steve Mao, Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn ASSOCIATE EDITORS Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Keith T. Smith (UK), Yury V. Suleymanov LETTERS EDITOR Jennifer Sills LEAD CONTENT PRODUCTION EDITORS Harry Jach, Lauren Kmec CONTENT PRODUCTION EDITORS Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Suzanne M. White SR. EDITORIAL COORDINATORS Carolyn Kyle, Beverly Shields EDITORIAL COORDINATORS Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Alana Warnke, Alice Whaley (UK), Anita Wynn PUBLICATIONS ASSISTANTS Jeremy Dow, Alexander Kief, Ronmel Navas, Hilary Stewart (UK), Brian White EXECUTIVE ASSISTANT Jessica Slater ASI DIRECTOR, OPERATIONS Janet Clements (UK) ASI SR. OFFICE ADMINISTRATOR Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis INTERNATIONAL EDITOR Martin Enserink DEPUTY NEWS EDITORS Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff SR. CORRESPONDENTS Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi ASSOCIATE EDITORS Jeffrey Brainard, Catherine Matacic NEWS REPORTERS Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad (Cambridge, UK), Paul Voosen, Meredith Wadman INTERN Alex Fox CONTRIBUTING CORRESPONDENTS Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Shanghai), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) CAREERS Donisha Adams, Rachel Bernstein (Editor), Katie Langin COPY EDITORS Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) ADMINISTRATIVE SUPPORT Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod GRAPHICS MANAGING EDITOR Alberto Cuadra PHOTOGRAPHY MANAGING EDITOR William Douthitt WEB CONTENT STRATEGY MANAGER Kara Estelle-Powers SENIOR DESIGNER Chrystal Smith DESIGNER Christina Aycock GRAPHICS EDITOR Nirja Desai INTERACTIVE GRAPHICS EDITOR Xing Liu SENIOR SCIENTIFIC ILLUSTRATORS Valerie Altounian, Chris Bickel SCIENTIFIC ILLUSTRATOR Alice Kitterman SENIOR GRAPHICS SPECIALISTS Holly Bishop, Nathalie Cary SENIOR PHOTO EDITOR Emily Petersen

Interim Chief Executive Officer and Executive Publisher Alan Leshner

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi DIRECTOR, BUSINESS OPERATIONS & ANALYSIS Eric Knott DIRECTOR OF ANALYTICS Enrique Gonzales MANAGER, BUSINESS OPERATIONS Jessica Tierney SENIOR BUSINESS ANALYST Cory Lipman, Meron Kebede FINANCIAL ANALYST Alexander Lee ADVERTISING SYSTEM ADMINISTRATOR Tina Burks SENIOR SALES COORDINATOR Shirley Young DIGITAL/PRINT STRATEGY MANAGER Jason Hillman QUALITY TECHNICAL MANAGER Marcus Spiegler ASSISTANT MANAGER DIGITAL/PRINT Rebecca Doshi SENIOR CONTENT SPECIALISTS Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy DIGITAL PRODUCTION MANAGER Lisa Stanford CONTENT SPECIALIST Kimberley Oster ADVERTISING PRODUCTION OPERATIONS MANAGER Deborah Tompkins DESIGNER, CUSTOM PUBLISHING Jeremy Huntsinger SR. TRAFFIC ASSOCIATE Christine Hall SPECIAL PROJECTS ASSOCIATE Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers GLOBAL MARKETING MANAGER Allison Pritchard DIGITAL MARKETING MANAGER Aimee Aponte MARKETING MANAGER Shawana Arnold MARKETING ASSOCIATES Tori Velasquez, Mike Romano, Ashley Hylton DIGITAL MARKETING SPECIALIST Asleigh Rojanavongse SENIOR DESIGNER Kim Huynh TRADE SHOW AND MEETINGS ASSOCIATE Andrew Clamp

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders ASSISTANT EDITOR, CUSTOM PUBLISHING Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT Sarah Whalen ASSOCIATE DIRECTOR, PRODUCT MANAGMENT Kris Bishop ASSOCIATE DIRECTOR, PRODUCT DEVELOPMENT AND SPJ Hannah Heckner SR. PRODUCT ASSOCIATE Robert Koepke DIGITAL PRODUCT STRATEGIST Michael Hardesty SPJ ASSOCIATE Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT Elisabeth Leonard MARKETING MANAGER Kess Knight SENIOR INSTITUTIONAL LICENSING MANAGER Ryan Rexroth INSTITUTIONAL LICENSING MANAGER Marco Castellan MANAGER, SYSTEMS AND OPERATIONS Brian Holiahn MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS Judy Lillibridge SENIOR OPERATIONS ANALYST Lana Guz FULFILLMENT COORDINATOR Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes US EAST COAST AND MID WEST SALES Stephanie O'Connor, Glen Cox US WEST COAST SALES Lynne Stickrod US SALES MANAGER, SCIENCE CAREERS Claudia Paulsen-Young US SALES REP, SCIENCE CAREERS Tracy Anderson ASSOCIATE DIRECTOR, ROW Roger Goncalves SALES REP, ROW Sarah Lelarge SALES ADMIN ASSISTANT, ROW Bryony Cousins DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA Xiaoying Chu ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION Grace Yao SALES MANAGER Danny Zhao PROJECT MANAGER Kilo Lan ASCA CORPORATION, JAPAN Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) COLLABORATION/CUSTOM PUBLICATIONS/JAPAN Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David RIGHTS AND LICENSING COORDINATOR Jessica Adams RIGHTS AND PERMISSIONS ASSOCIATE Elizabeth Sandler CONTRACTS AND LICENSING ASSOCIATE Lili Catlett

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
sciencemag.org/authors/science-information-authors

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

EDITORIAL

Getting the EPA back on track

Your information will be kept confidential, and the lessons learned from your participation will serve society—those are the promises made by researchers to participants in studies designed to inform environmental policies, from clean water and air to chemical exposure limits. The United States Environmental Protection Agency (EPA) may well break this fundamental pact next year, putting the agency at odds with its very mission "to protect human health and the environment." Hopefully, the EPA will realize that this would jeopardize regulations that keep the environment safe to live in, and correct course back to sound policy-making.

In January 2020, the EPA plans to issue a supplement to its 2018 proposed rule, Strengthening Transparency in Regulatory Science, which stated that in setting standards, the agency would only use research for which underlying raw data and models were made public. The rule could eliminate many public health studies from consideration. At a congressional hearing last month, the EPA claimed that the supplemental rule provides clarifications, but does it address major problems with the plan? Although the notion of depositing data and models from federally funded research into public databases is laudable, the rule as proposed poses substantial problems. This may account for why the majority of nearly 600,000 public responses to the 2018 proposed rule were critical.

> **"The EPA's proposed transparency rule… unquestionably excludes key science from policy-making."**

In epidemiological and clinical studies, people provide information—their medical histories, behaviors, education, employment, and other personal details—under the condition that it will not be shared and their privacy will be protected. Anonymizing data is already difficult, if not impossible. With geographically referenced data, a capable programmer can leverage machine learning and brute computational strength to determine the location, and subsequently the identity, of a study participant. Similarly, facial recognition software has been applied to images reconstructed from cranial scans to identify study participants. Reidentification can jeopardize employment, insurance, or personal relationships for individuals, and scholarship, reputation, or funding for researchers. This will simply discourage people from participating in future health studies. Moreover, successfully recruiting and retaining participants depends on trusting relationships built on meaningful and sustained interaction between researchers and participants, especially with disadvantaged populations who are underrepresented in research. The EPA rule assumes that people will consent to their data residing in a repository where decisions about data use are made by persons unknown to them.

The proposed rule claims that additional analysis of raw data and models will improve science. Who will do this analysis? Most likely, vested interests will finance work slanted toward a particular outcome, rather than undertake scientific inquiry without an agenda. For example, lead paint industry defense attorneys have attributed children's neurological deficits to landlord neglect and parental failure. The rule also disregards the power of the "weight of the evidence." Imagine multiple studies done by different investigators on different populations using different techniques, yet reaching similar conclusions—that's a powerful result. Ignoring the weight of evidence derived from the totality of relevant science, regardless of data availability, contravenes the EPA's directive (stated in the Clean Air Act) to set standards "requisite to protect the public health" with "an adequate margin of safety."

Many researchers already deposit code and data into open repositories. The U.S. National Institutes of Health and other federal funding agencies require data-sharing plans to support independent reanalysis within the scientific community without compromising confidentiality. The peer review process provides an additional check on the credibility of research results. Work by the Health Effects Institute, in which an industry-government–funded partnership reanalyzed data from the Harvard Six Cities Study and the American Cancer Society Study on the link between particulate matter pollution and mortality, represents an excellent model for evaluating the validity of research pivotal to environmental health regulations without compromising confidentiality or excluding studies.

The EPA's proposed transparency rule does not ensure research rigor or improve transparency. It unquestionably excludes key science from policy-making. Once the supplemental rule is released in January 2020, there will be an open period for public comment—an opportunity for everyone to remind the EPA of its obligation to use the best science, as required in multiple environmental laws, to protect human health and the environment.

–Marie Lynn Miranda

Marie Lynn Miranda is a professor in the Department of Statistics at Rice University, Houston, TX, USA, and is the founding director of the Children's Environmental Health Initiative at Rice University. mlm@rice.edu

PHOTO: THOMAS LAVERGNE

10.1126/science.aba3769

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

Record high temperatures across Europe this summer sent people wading in fountains in Paris.

CLIMATE SCIENCE

Older climate models pass temperature test

Older climate models, even those created half a century ago, faithfully predict recorded increases in temperature, a study has found—suggesting scientists can trust model predictions of future warming caused by rising levels of greenhouse gases. Critics of climate science often claim such models are flawed. But the new analysis determined that the predictions of 10 out of 17 past climate models released between 1970 and 2001 tracked with measurements of global temperatures. And four of the inaccurate models improved when scientists accounted for unanticipated changes in the amount of greenhouse gases humans generated, they report in the 4 December issue of *Geophysical Research Letters*. Difficult-to-predict influences such as new regulations can skew emissions and throw off a model's temperature predictions, explains the paper's lead author, Zeke Hausfather of the University of California, Berkeley. But the overall accuracy of even older models, simpler than those used today, confirms that scientists' understanding of the physics behind greenhouse warming is sound, he adds. The findings come as officials from around the world meet this week in Madrid for the United Nations's annual conference on climate change, with a focus on implementing pledges to reduce greenhouse gases.

FDA approves sickle cell drug

DRUG DEVELOPMENT | The U.S. Food and Drug Administration (FDA) on 25 November approved a novel therapy for sickle cell disease, an often fatal blood disorder that afflicts about 100,000 Americans. The drug, developed by Global Blood Therapeutics based in South San Francisco, California, is the first to target the disease's root cause—a mutant form of the protein hemoglobin, which carries oxygen inside red blood cells. Critics of the drug complain that in clinical trials it improved people's blood test results but not symptoms, such as pain caused when sickled cells block small blood vessels. FDA may approve drugs for serious conditions based on improvements in "surrogate" endpoints such as blood tests if other treatments are lacking. The sickle cell drug will cost about $125,000 annually; the firm said it will explore how to provide the treatment affordably in developing countries.

PHOTO: SAMUEL BOIVIN/NURPHOTO VIA GETTY IMAGES

Society policy targets harassers

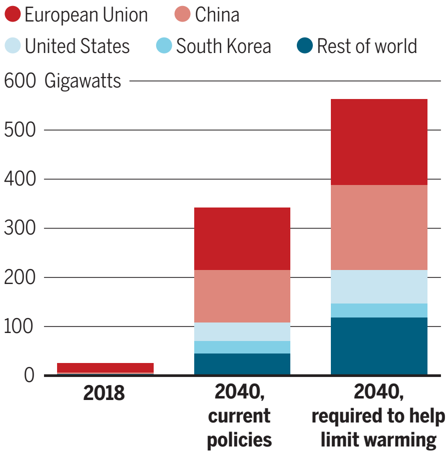
SCIENTIFIC SOCIETIES | Members of the Society for American Archaeology (SAA) have voted to allow its board to bar attendees who have committed sexual harassment or other misconduct from society events. The vote to change SAA's bylaws, announced on 22 November, was a response to a scandal sparked when an archaeologist banned from his university for sexual harassment was allowed to attend the society's annual meeting in April. The new policy, written by SAA's board, also covers bullying and other forms misconduct. But some SAA members are concerned the wording is not strong enough because it states the society's board "may" bar offenders from events; they had proposed an alternate bylaws amendment that offenders "will be barred." That measure failed in the vote.

Off-shore wind power grows

ENERGY | By 2040, off-shore wind turbines could produce about as much electricity as those onshore and, with solar cells, supply much of the renewable energy required to reduce global warming, says a forecast published last month. To date, more wind turbines have been built on land, but turbines placed in coastal waters offer the advantage of steadier winds, and costs could become competitive with fossil fuels in the next decade, says the International Energy Agency's *World Energy Outlook 2019*. The European Union and China will dominate off-shore wind power through 2040, the report says. But it also warns that current growth in wind and other renewables won't reduce greenhouse gas emissions by enough to limit the increase in global temperatures to less than 2°C; even larger growth in wind and other alternative energies is needed.

▲ Mighty wind

Off-shore wind power is growing quickly—but needs to expand faster still to help control global warming.



CREDITS: (GRAPHIC) J. BRAINARD/*SCIENCE*; (DATA) INTERNATIONAL ENERGY AGENCY; *WORLD ENERGY OUTLOOK 2019*

Disgraced surgeon sentenced

MISCONDUCT | A court in Italy has sentenced surgeon Paolo Macchiarini to 16 months in prison for forgery and abuse of his office. Macchiarini became famous a decade ago while at the Karolinska Institute in Stockholm for transplanting artificial windpipes "seeded" with stem cells into patients. The transplants failed and nearly all the patients died; Macchiarini was found guilty of scientific misconduct and charged with manslaughter in Sweden. Those charges were dropped in October 2017, but now, an old case in Italy has apparently caught up with him. In 2012, Macchiarini was charged with taking bribes from patients. After years of appeals and counter appeals, on 8 November a court found him guilty of minor infractions: providing treatment for a friend who was not eligible for care and manipulating the patient's records. Macchiarini's lawyer said he will appeal the sentence. An Italian newspaper reports the surgeon is now working in Japan.

U.K. pushes for AI accountability

COMPUTER SCIENCE | Businesses and government offices that use artificial intelligence (AI) to help make decisions—to predict health outcomes, extend credit, or even determine criminal penalties—will soon be required to explain how their technology functions to individuals affected by it, according to draft guidance from the United Kingdom's Information Commissioner's Office this week. The guidance, thought to be the world's first, is meant to "get out ahead" of potential abuses, says David Leslie, an ethics fellow at the Alan Turing Institute in London and a co-author of the draft. The explanations of the software's methods and reliability must be comprehensible by lay people, who could use existing data-privacy or antidiscrimination laws to challenge how the AI is applied. The office is accepting comments on the 165-page draft until 24 January 2020.

Elsevier inks open-access deals

SCHOLARSHIP | Publishing giant Elsevier last month signed two new "transformative" deals allowing researchers to read its journals and publish in subscription ones on an open-access basis. Its deal with Carnegie Mellon University in Pittsburgh, Pennsylvania, is Elsevier's first of this kind with a U.S. institution. Its other deal is with the Bibsam Consortium, which negotiates such agreements for Sweden's universities; they gave up subscriptions to Elsevier journals in 2018 after Bibsam unsuccessfully sought a new, open-access agreement. Carnegie Mellon did not disclose the cost of its deal; Bibsam said its outlay will depend on uptake by its library members.

BY THE NUMBERS

48

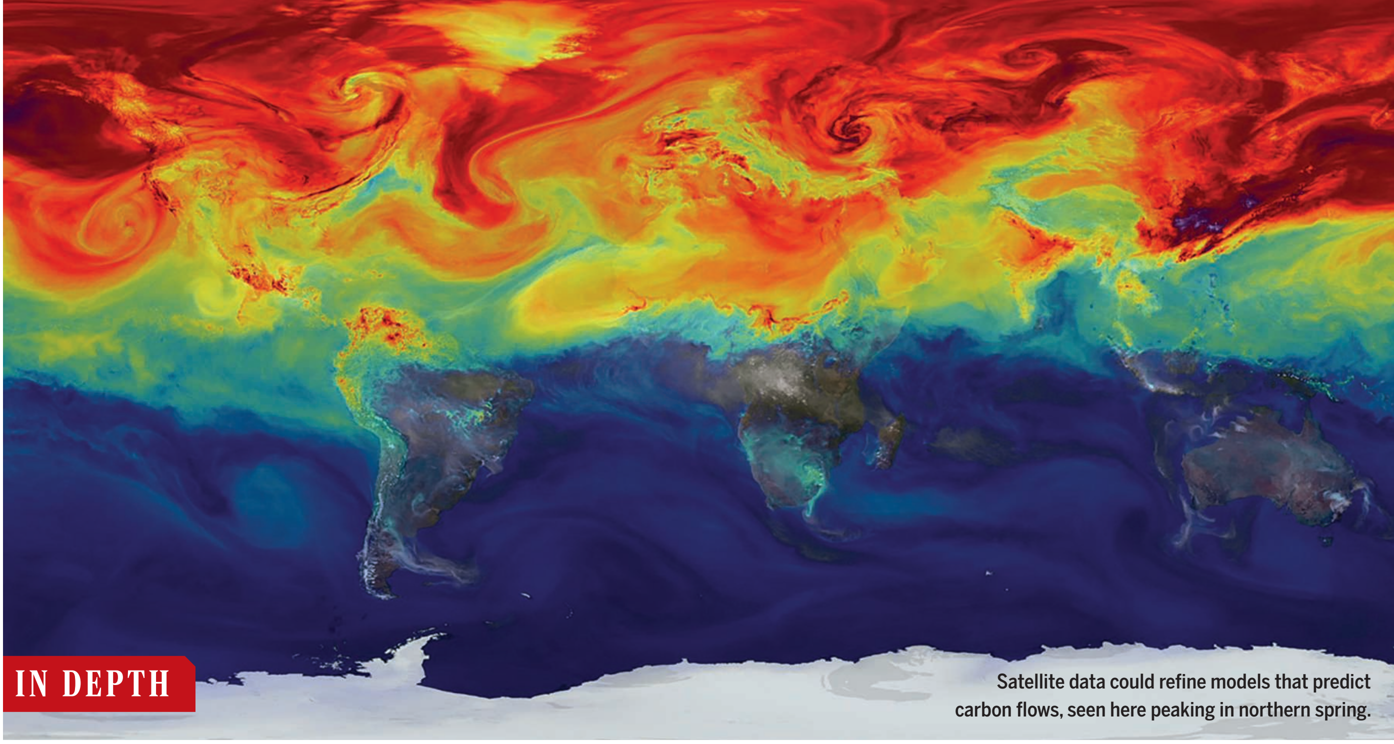
U.S. states in which mortality rates among 25- to 64-year-olds increased from 2010 to 2017(*JAMA*).

1.6 million

Hectares in Indonesia that burned this year, according to Sentinel satellite data, an area half the size of Belgium. That figure is twice the Indonesian government's estimate, based on lower resolution LANDSAT data (Center for International Forestry Research).

Arab scientists want to emigrate

WORKPLACE | Although many Arab countries want to develop "knowledge economies" to diversify their petroleum-based industries, 91% of their researchers would like to emigrate for professional opportunities elsewhere, a survey has found. The findings come from an online poll of 650 researchers working in 22 countries of the Arab League, reported this week by Al-Fanar Media, a nonprofit news organization that covers research and education. Of the survey respondents, 232 work in science, technology, engineering, and math (STEM) fields, most of them at universities. Poor funding was a frequent complaint about their current jobs, but so was the professional environment. Among STEM workers, 87% of those wanting to move desired to improve their research. Other incentives included better research facilities (67%), more salary (43%), and more academic freedom (40%).



Satellite data could refine models that predict carbon flows, seen here peaking in northern spring.

SPACE SCIENCE

Europe to lead in monitoring carbon from space

Budget hike for ESA's Copernicus program advances satellites to monitor Paris accord cuts

By **Daniel Clery**

Even optimists at the European Space Agency (ESA) were startled last week when its member governments awarded it a €12.5 billion, 3-year budget, its largest ever and more than 20% above its previous 3 years of funding. With the unexpected windfall, ESA will develop a reusable space cargo capsule, support the International Space Station until 2030, and join NASA in retrieving rocks from Mars.

But one of the biggest winners, up 29% to €1.8 billion, is Copernicus, a program supporting a fleet of satellites that continuously tracks features of Earth's atmosphere and surface, including the contours of the sea surface and shifts in vegetation. The money will help Europe expand the fleet to observe humanmade sources of carbon dioxide (CO_2) on a daily basis—making ESA the only space agency capable of monitoring pledges made under the Paris accord to cut greenhouse gases. Europe's CO_2 monitoring plans are "unparalleled," says Christopher O'Dell, an atmospheric scientist at Colorado State University in Fort Collins. "The Europeans are just running with this."

Just why ministers from ESA's 22 member states were feeling so generous at a meeting last week in Seville, Spain, is unclear. It could be because Europe's economy is better than 3 years ago, or because ESA officials did a good job talking up plans with ministers and stakeholders, says Athena Coustenis, a planetary scientist at the Paris Observatory and chair of the European Space Sciences Committee, an advisory body. Regardless, the only mission not to receive full funding was Lagrange, a set of space weather satellites. "There's no need to kill anything," Coustenis says. "I'm in shock."

ESA's science budget, stagnant for decades, got a 10% hike to €2.8 billion. That should allow the agency to study black holes with two concurrent missions, a gravitational wave detector and an x-ray observatory (*Science*, 25 October, p. 410). The exploration budget was boosted by one-third, which means ESA can launch the ExoMars rover next year—if it can fix problems with its parachutes in time (*Science*, 29 November, p. 1061). The money will help ESA join NASA's Artemis program to build a Moon-orbiting space station called the Gateway and work toward a human presence on the surface. It will also pay for initial work on a complex mission to bring back samples from Mars. NASA hopes to follow suit next year (*Science*, 22 November, p. 932).

The Copernicus system of Earth-observing Sentinel satellites also got a lot of love. This joint venture with the European Union provides long-lived, unbroken data sets to government, industry, and academic users. Existing Sentinel satellites monitor, for instance, land use and sea surface height. The first three Sentinels are now operational, along with duplicate satellites that serve as backups. Three more Sentinels are in the works. When it comes to Earth observation, "Europe has the most capable fleet in orbit," says Martin Visbeck of the GEOMAR-Helmholtz Centre for Ocean Research in Kiel, Germany, and chair of ESA's Advisory Committee for Earth Observation.

The funding boost ensures ESA can proceed with ambitious plans for a further set of six candidate Sentinels, among them the CO_2 monitoring mission. Around the time of the Paris climate pact of 2016, the European Union decided that national carbon budgets, based on disclosures from known emitters such as power plants and cement works, needed checking from the sky. In response, ESA resurrected plans for CarbonSat, which failed to win a launch slot in 2015, and beefed it up into a Sentinel, which would look for the spectral absorption signals of CO_2 in infrared sunlight reflected off Earth's surface.

With the new funding, a CO_2 Sentinel could launch as soon as 2025, putting Europe in position to contribute to a census of emissions that the Paris accord says should take place every 5 years beginning in 2023. Carbon dioxide is "a quantity we need to watch for years to come," Visbeck says.

The ministers gathered in Seville appar-

IMAGE: NASA GODDARD SPACE FLIGHT CENTER

Published by AAAS

ently shared that sense of urgency. "Governments are really putting words into action," says Josef Aschbacher, ESA's director of Earth observation in Frascati, Italy. But their motivations may not be entirely selfless, because the contracts for Copernicus are spread among member states in line with their contributions. "It's a unique opportunity to position your national industry for decades to come," Aschbacher says.

NASA pioneered efforts to track CO_2 from space with its Orbiting Carbon Observatory-2 (OCO-2), launched in 2014. But OCO-2 samples a narrow swath, returning to each point on Earth at intervals of weeks, and was only designed to last 2 years. China's current CO_2-tracking satellite TanSat, which closely follows OCO-2's design, has struggled with its calibration. And OCO-3, launched this year and attached to the space station, is another short-lived research mission; the United States has no plans to follow up with an operational carbon-monitoring system. "It's a little frustrating watching from the U.S.," O'Dell says. "We've lost the leadership role in this."

ESA's CO_2 mission would operate for 20 years, with as many as three identical satellites scanning the entire globe with swaths 300 kilometers wide. The mission would scan each point on Earth every few days, capturing the changing plumes of individual power plants. "We're moving from 1D to 2D," says Michael Buchwitz, an adviser to the project at the University of Bremen in Germany. In a step up from the original CarbonSat proposal, the CO_2 Sentinels will also help identify emitters by detecting nitrogen oxides from fossil fuel combustion, and will have sensors for clouds and aerosols to improve accuracy.

The European Union's next 7-year budget, now being negotiated, could determine whether the mission reaches its full potential. ESA is only responsible for the first Sentinel of each type, with the EU paying for the backups. And as the user of the data, the EU will also finance a network of ground sensors to calibrate the satellite carbon measurements as well as data-processing and modeling efforts. "The satellites are a core component," Buchwitz says. But there is also "much, much more."

If Europe sustains its carbon-monitoring efforts, they will be a boon to scientists as well as policymakers, O'Dell says. The high-resolution Sentinel readings will help scientists tune their models of how CO_2 flows around the atmosphere. And the missions could also attract more scientists to what remains a small field, he adds. "It will bring more brain power." ■

With additional reporting by Paul Voosen.

PUBLIC HEALTH

Beset by neural tube defects, Ethiopia may fortify salt

Government wants evidence of effectiveness before adding folic acid to ubiquitous dietary staple

By **Meredith Wadman**

Tony Magana, chief of neurosurgery at Mekelle University School of Medicine in Ethiopia's Tigray province, is constantly reminded of his country's high prevalence of neural tube defects (NTDs). His team operates on more than 400 babies annually to repair these severe, often lethal birth malformations, in which babies can be born without brains or with their spinal cords protruding from their backs. "Probably every other day we see a child that is so bad we can't help them," Magana says.

Last month, a team of nutrition experts converged in Addis Ababa to lay groundwork for an unproven but possibly highly effective intervention: fortifying Ethiopia's salt supply with folic acid, a synthetic form of the B vitamin folate. In the first 4 weeks of pregnancy, folate is essential to proper closure of the neural tube, which gives rise to the brain and spinal cord, and since the mid-1990s, more than 80 countries have mandated flour fortification with folic acid. Ethiopia, where fewer than one-third of people eat flour, is not among them.

Last year, studies that surveyed births at 11 hospitals there shook the global health community. The studies—one co-authored by Magana—found that among every 10,000 births, between 126 and 131 babies had NTDs. That's seven times their global prevalence and 26 times that of some high-income, flour-fortifying countries. According to Ethiopian government data, 84% of Ethiopian women of reproductive age have folate levels in their red blood cells low enough to put them at risk of having a child with an NTD.

"These numbers from Ethiopia are some of the worst anywhere and ever," says Marinus Koning, a retired surgeon who is founder of the ReachAnother Foundation, a charity based in Bend, Oregon, that has supported the training of dozens of Ethiopian neurosurgeons. "Something needs to be done about it."

At the invitation of the Ethiopian Ministry of Health, Koning and scientists from the United States, Canada, and the Netherlands began to work with experts at the Ethiopian Public Health Institute (EPHI) in Addis Ababa to develop a plan to address NTDs. The result was an issue brief released by EPHI in May that recommended the government consider salt fortification.

The prospect is winning praise from affected families. "We need prevention more than any intervention," says Beza Haile, founder of the Addis Ababa–based advocacy group HOPE-Spina Bifida and Hydrocephalus. Haile's 4-year-old son, Hezkiel, who has an NTD, can't talk, walk, sit, or eat, except foods that are the consistency of soft porridge.

A method of fortifying salt—spraying it with buffered folic acid solution—had already been developed by chemical engineer

PHOTO: REACHANOTHER

Babies recover from surgery for spina bifida, a malformation of the spine and spinal cord, in Hawassa, Ethiopia.

Levente Diosady and colleagues at the University of Toronto in Canada. The same spraying equipment used for iodization of salt, already mandated in Ethiopia to prevent intellectual disabilities and thyroid disease, can deliver folic acid. "One of the main reasons this project is moving forward and there is a lot of political support for it is it requires few adaptations," says Christine McDonald, a micronutrient scientist at Children's Hospital Oakland Research Institute in California.

McDonald was part of the team that visited Addis Ababa last month. There, the team met with potential funders and with Hakan Kolenoğlu, chief executive of the country's leading salt processor, SVS Salt Production PLC, who promised to fortify, free of charge, 40 tons of salt for preliminary studies.

With initial funding from ReachAnother (more will be needed, the team says), researchers will test whether folate-fortified salt is stable in Ethiopian environmental conditions and whether its sensory qualities, including a yellowish tinge, are acceptable to Ethiopians. If the answers are encouraging, fortified salt's effects on the gold standard measurement of folate sufficiency—red blood cell folate levels—will be put to the test in a randomized, controlled, double-blind trial of hundreds of women of reproductive age.

"There is no scientific evidence that adding [folic acid] to salt could improve the folate status of women," says Masresha Tessema, a nutritionist at EPHI's Food Science and Nutrition Research Directorate who was first author on the issue brief and is the Ethiopian lead on the planned studies. "The ministry needs evidence."

If salt supplementation works, it could be game changing for Ethiopia: A meta-analysis this year concluded that large-scale folic acid food fortification in low- and middle-income countries has lowered the risk of NTDs by 41%. "We have an amazing opportunity to do a lot of good," says Kenneth Brown, the lead U.S. scientist on the team that met in Addis Ababa. Brown, an emeritus professor at the University of California, Davis, who was until recently a senior nutrition scientist at the Bill & Melinda Gates Foundation, adds: "It's shovel ready. We know what the problem is. We know how to fix it."

Other experts hope an Ethiopian success story could spur efforts in more than 110 other countries that don't mandate food fortification. Says Nicholas Wald, an epidemiologist at University College London, who in a seminal 1991 paper established that taking 4 milligrams of folic acid daily, before and in early pregnancy, reduces the risk of NTDs by about 80%: "It's a global issue of which Ethiopia is an extreme example. Loads of countries should be fortifying a staple food with folic acid and aren't." ■

A new Italian funding agency would be supervised by the Ministry of Education, University, and Research in Rome.

PHOTO: VITO ARCOMANO/ALAMY STOCK PHOTO

EUROPE

Italy set to create €300 million research funding agency

Researchers applaud plan but worry about agency's small budget and potential political interference

By **Giorgia Guglielmi**

Scientists in Italy are about to receive a long-sought gift—but some are disappointed. This month, Italy is expected to set up its first national science funding agency, with an annual budget that would rise to €300 million. Italian scientists are welcoming the boost to a thin basic research budget and the prospect of an independent body that could allocate the money transparently. But some complain that the sum is too small and worry that the new National Research Agency (ANR) will be vulnerable to political interference.

Originally announced in September by Prime Minister Giuseppe Conte, who leads a coalition government of the populist Five Star Movement and the center-left Democratic Party, the ANR proposal is now a part of the country's 2020 budget bill, which the Parliament must approve by 31 December. The Senate is set to vote on the bill this week, after which it will move to the lower house. The agency could be up and running in a matter of months—with many questions hanging over it.

"There is the willingness to set up a national science funding agency, but there is not much clarity on how it should be done," says Maria Cristina Messa, a clinical diagnostics researcher at the University of Milan-Bicocca. Messa adds that €300 million might be adequate to fund basic research projects. "But for applied research, it's definitely not enough," she says.

With R&D spending of about 1.3% of its gross domestic product, Italy lags behind the 2.4% average of other developed countries. Nearly one-third of the spending, which totaled €23.8 billion in 2017, according to the Italian National Institute of Statistics, came from public sources such as the Ministry of Health, the Ministry of the Environment, and the Ministry of Education, University, and Research (MIUR). But ministry funding can be erratic and lacks long-term planning, says Piergiuseppe De Berardinis, an immunologist at the Italian National Research Council in Naples. For example, MIUR's last grant call for basic research was in 2017, and grant winners began to receive the money only this year. What's more, grant review is often a black box, De Berardinis says. Italian public agencies approve or reject applications "with just a few lines of boilerplate text, which leaves no room for appeal," he says.

ANR won't solve these problems, as it will be an add-on to the current system rather than a replacement, says Nicola Bellomo, a mathematician at the Polytechnic University of Turin and president of Gruppo 2003, a group that has called for a national funding agency. But it could make some annual

grant calls and distribute its limited funds efficiently and transparently, Bellomo says.

The current bill says ANR's annual budget would start with €25 million in 2020, and rise to €200 million in 2021 and €300 million in 2022. It would fund "innovative" projects, help coordinate research at universities and public agencies, and promote Italy's participation in international projects.

The bill also says ANR will be an independent agency working "under the supervision" of the prime minister and the research minister. Most appointments to the agency would be political: The prime minister would select the director, and the politically appointed ministers of several government agencies would appoint five of the eight members of the ANR executive board. "That's very troubling," says Alberto Baccini, an economist at the University of Siena, who worries that political appointees could steer funds to fit an ideological agenda. To safeguard the agency's political independence, Baccini says scientists should elect at least some of its management. Others hope an international search committee would select experts to ANR's executive board.

The head of MIUR's technical secretariat, Fulvio Esposito, who helped write the draft bill, admits that the current proposal is a hastily written outline for the new agency. Esposito says MIUR is trying to get the Parliament to simplify the bill so that it establishes the agency without specifying how it will be staffed and run. That would allow MIUR officials to spend time studying the structures and funding models of research agencies in other developed countries and gather feedback from Italian researchers.

James Wilsdon, a science policy researcher at the University of Sheffield in the United Kingdom, cautions that adding another funding organization, which could take years to set up completely, might not be the best way to boost Italian research. Wilsdon also worries that friction about ANR's structure could become a distraction from increasing Italy's overall research spending, "which is clearly what's required." Although €300 million won't propel Italy past other nations, he says it's a nontrivial amount, and that it's sensible to start small, see what works, and scale up.

Others are more optimistic. The creation of ANR signals an ambition to turn around the fortunes of Italian research, says molecular biologist Rosario Rizzuto, president of the University of Padua. If adequately structured and funded, Rizzuto says, the agency can be "a good opportunity to boost investments that are necessary for the economic growth of our country." ■

Giorgia Guglielmi is a journalist in Cambridge, Massachusetts.

CREDITS: (GRAPHIC) A. CUADRA/*SCIENCE*; (DATA) PCORI

MEDICINE

Institute that aims to reshape health care seeks renewal

Mixed reviews for effort to compare medical treatments

By **Jocelyn Kaiser**

Patients and their doctors face crucial choices every day: Surgically excise precancerous breast cells or watch carefully for growth. Try meditation for anxiety or go with prescription drugs. Stay in the hospital for tests after chest pain or head home and have testing later. To help with such decisions, and to rein in U.S. health care costs, Congress 9 years ago created an independent research institute that would enlist patients as partners in designing studies that compare the benefits of established medical treatments.

Now, after awarding nearly $2.6 billion for research, the nonprofit Patient-Centered Outcomes Research Institute (PCORI) in Washington, D.C., faces a turning point. The agency's charter was set to expire on 30 September, and it is operating under a temporary extension until 20 December. Congress must decide whether to reauthorize the organization long term, and if it does, whether to fine-tune its mission.

Supporters say it's obvious that PCORI should continue. With annual incoming funds averaging about $480 million since 2014, the operation has paid for more than 700 projects, more than half completed by now, that are already influencing health care. "We're starting to see results that really matter," says Christine Goertz, a health services researcher at Duke University in Durham, North Carolina, who chairs PCORI's governing board.

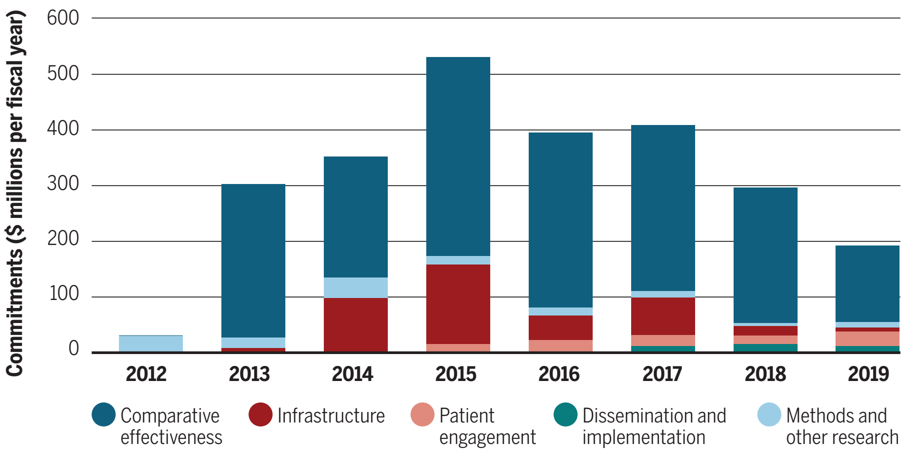
In spite of complaints that PCORI was slow to launch large randomized trials that could sway medical practice and cut costs, disease advocacy and research groups back it. "They have picked up speed and set up more effective studies. What they're doing now is what they should be doing," says Ross McKinney, chief scientific officer for the Association of American Medical Colleges, based in Washington, D.C.

Lawmakers seem to agree—bipartisan bills in the House of Representatives and Senate would reauthorize PCORI for as long as 10 years. But with a crowded legislative calendar and impeachment taking lawmakers' attention, it's not clear when a bill will be passed.

PCORI's birth took political compromises. Congress created it as part of then-President Barack Obama's 2010 Affordable Care Act, funding it largely from a tax on health insurance plans that goes into a trust fund. Backers reasoned that "comparative effectiveness research" could help control ballooning medical costs. Conservatives worried, however, that PCORI's work would result in health care rationing, so

Spend a little, save a lot?

By funding studies that compare treatments, the Patient-Centered Outcomes Research Institute promises to control health care costs.



lawmakers barred the agency from making cost comparisons while mandating that its board include stakeholders representing industry, insurers, patients, and physicians. They also gave it a name that did not mention comparative research, but instead reflected a less controversial goal: putting the needs and interests of patients first.

Figuring out what "patient-centered research" meant took time. "Nobody really knew what it was or used [the term] in their own way. In fact, PCORI defined it," says Ellen Sigal, chair of Friends of Cancer Research, a Washington, D.C.–based patient group. PCORI-funded researchers must try to include patients every step of the way, from enlisting them in study design to, at times, including them on the research team; patients also help PCORI review research proposals. The requirements "have been an adjustment" for many scientists, but worthwhile overall, says health services researcher Michael Fischer of the Harvard University–affiliated Brigham & Women's Hospital in Boston, who has worked on PCORI projects.

PCORI says patients have suggested changes to studies, such as emphasizing the ability to live at home as a treatment outcome and changing how a study is explained to them. PCORI has inspired U.S. agencies, drug companies, and health care organizations to incorporate patient-centeredness into their work, says Marc Boutin, CEO of the National Health Council in Washington, D.C., which represents patient groups.

But developing methods for involving patients and setting up review panels was time consuming, and PCORI's first contracts only went out in late 2012. Many early projects focused on patient education and engagement, such as evaluating "decision tools," or brief information sheets, to help people and their caregivers choose between treatments. The institute also spent time and money to build PCORnet, a clinical research network that pools electronic health records for millions, which aids in comparisons of treatments and recruitment of patients for clinical trials.

PCORI eventually ramped up its support of comparative effectiveness research, which has consumed 73% of its total award commitments to date, the institute says. Published findings include that oral antibiotics are as helpful as intravenous ones for certain children with serious infections, and that type 2 diabetics who aren't on insulin don't need to test their blood sugar every day. Some results have found their way into practice guidelines, and PCORI is now funding efforts to get physicians and health systems to adopt its results.

There has been "substantial disappointment" in the slow pace, however, says bioethicist Ezekiel Emanuel of the University of Pennsylvania, who helped craft the PCORI authorization as an Obama official. By now, he adds, it should be able to "rattle off five or 10 things" that have had a major impact on medical practice. Emanuel contends that its patient engagement efforts went "way overboard." He also faults the institute for shying away from large-scale drug comparisons for fear of "antagonizing" the drug companies. "They weren't bold enough," Emanuel says.

Health insurers, too, feel that PCORI "hasn't really informed our decision-making as of yet," says Kate Berry, senior vice president for strategic partnerships for America's Health Insurance Plans, a lobbying group in Washington, D.C.

PCORI's defenders say its critics have "an unrealistic understanding of how rapidly you could implement large scale trials. Then there's the lag at the other end—how sluggish the adoption in health care is," says PCORI interim Executive Director Josephine Briggs, a former National Institutes of Health (NIH) official who replaced its founding director Joe Selby last month. She also defends its decision not to fund more of the long-term drug comparison studies more typically done by NIH. Their cost "could absorb our budget and keep PCORI from doing more of the very practical stuff," she says.

> **"They have picked up speed and set up more effective studies."**
> **Ross McKinney**, Association of American Medical Colleges

Although PCORI still has funds left to administer ongoing grants and make a few additional awards, it has no money to initiate new rounds of funding. A search for a permanent director is underway, but Briggs expects any candidate will want to wait to see whether reauthorization goes through.

Two bills approved by House panels would renew the institute for only three to seven more years and leave its operations largely unchanged. A bill introduced in the Senate would give PCORI another 10 years, creates an advisory panel to identify high-impact research areas, requires more short-term projects, and allows the institute to consider treatment costs—all changes that Berry's group and other insurers favor.

Although PCORI's supporters expect a compromise bill to pass in the coming months, they are watching the process anxiously. After nearly a decade spent building a new patient-centered way to fund health outcomes research, "it would be such a pity if this were to dwindle away," Fischer says. ■

ATMOSPHERIC CHEMISTRY

Nitrogen crisis threatens Dutch environment—and economy

Ecological damage from manure fumes triggers calls for drastic change to livestock industry

By **Erik Stokstad**

Last week, Dutch farmers across the country parked their tractors along highways in the third such protest since October, when they jammed traffic while driving en masse to The Hague, the nation's center of government.

They are protesting a Dutch high court decision that in May suspended permits for construction projects that pollute the atmosphere with nitrogen compounds and harm nature reserves. The freeze has stalled the expansion of dairy, pig, and poultry farms—major sources of nitrogen in the form of ammonia from animal waste. Also blocked are plans for new homes, roads, and airport runways, because construction machinery emits nitrogen oxides. All told, the shutdown puts some €14 billion worth of projects in jeopardy, according to ABN AMRO Bank. "It has really paralyzed the country," says Jeroen Candel, a political scientist at Wageningen University & Research.

The government is preparing to enact short-term measures, including lowering a highway speed limit, which could reduce nitrogen emissions a sliver. To make a significant dent, many experts say the country's farm animal sector—the densest in the world—must shrink and recycle more of its nitrogen. Last month, farmers asked for nearly €3 billion over 5 years to help pay for more environmentally friendly ways to deal with manure. Although agriculture is a large source of nitrogen emissions, other sectors will have to rein in their pollution, too. "Those are really tough political decisions that have to be made," says Jan Willem Erisman, a nitrogen expert at the Louis Bolk Institute in Bunnik.

Nitrogen, a key nutrient for plants, is also an insidious pollutant. Fertilizer wash-

Dutch farmers have protested a ruling that curtails the expansion of livestock operations because of the nitrogen pollution they produce.

PHOTO: VINCENT JANNINK/ANP/AFP/GETTY IMAGES

ing off fields ends up in lakes and coastal areas, causing algal blooms that kill marine life. Airborne nitrogen can also harm ecosystems. One source is nitrogen oxides, mostly from power plants and engine exhaust. In the Netherlands, even more comes from the ammonia vapors from livestock urine and manure. Both kinds of nitrogen react to form aerosols that cause smog, damage foliage, and acidify the soil, hindering roots' absorption of nutrients. (Dutch farmers must add lime to their fields to fight acidity.)

The Netherlands is a nitrogen hot spot partly because it is a dense, urbanized nation, although controls on power plants and catalytic converters in autos have helped curb nitrogen oxide emissions. The bigger problem is ammonia emissions from concentrated livestock operations. Dutch farms contain four times more animals biomass per hectare than the EU average. Practices such as injecting liquid manure in the soil and installing air scrubbers on pig and poultry facilities have reduced ammonia emissions 60% since the 1980s, but they have risen slightly since 2014 because of expanding dairy operations. Dutch agriculture is responsible for nearly half of nitrogen pollution that falls in the country.

In 118 of 162 Dutch nature reserves, nitrogen deposits now exceed ecological risk thresholds by an average of 50%. In dunes, bogs, and heathlands, home to species adapted to a lack of nitrogen, plant diversity has decreased as nitrogen-loving grasses, shrubs, and trees move in. Heathlands are turning green-gray as invasive grasses overwhelm the purple heather and yellows and blues of small herbaceous flowering plants, says Eva Remke, an ecologist at B-WARE Research Centre in Nijmegen. "The grasses will win, and the herbs will lose." These losses cascade through the ecosystem, contributing to the decline of insect and bird diversity, she says.

To control emissions, in 2015 the Netherlands introduced a nitrogen permit system that allows construction if, for example, regional governments reduce nitrogen from other sectors, such as farming. The system relies on a model developed by the Dutch National Institute for Public Health and the Environment (RIVM) to calculate how much nitrogen is emitted by various activities and how much they contribute to pollution in natural areas.

The system was not enough to satisfy environmental groups. They sued the Dutch government in 2016, demanding that it deny construction permits for expanded animal operations near two nature reserves. The cases ended up in the Court of Justice of the European Union, which last year ruled against the government and criticized the permit system for not ensuring immediate nitrogen reductions.

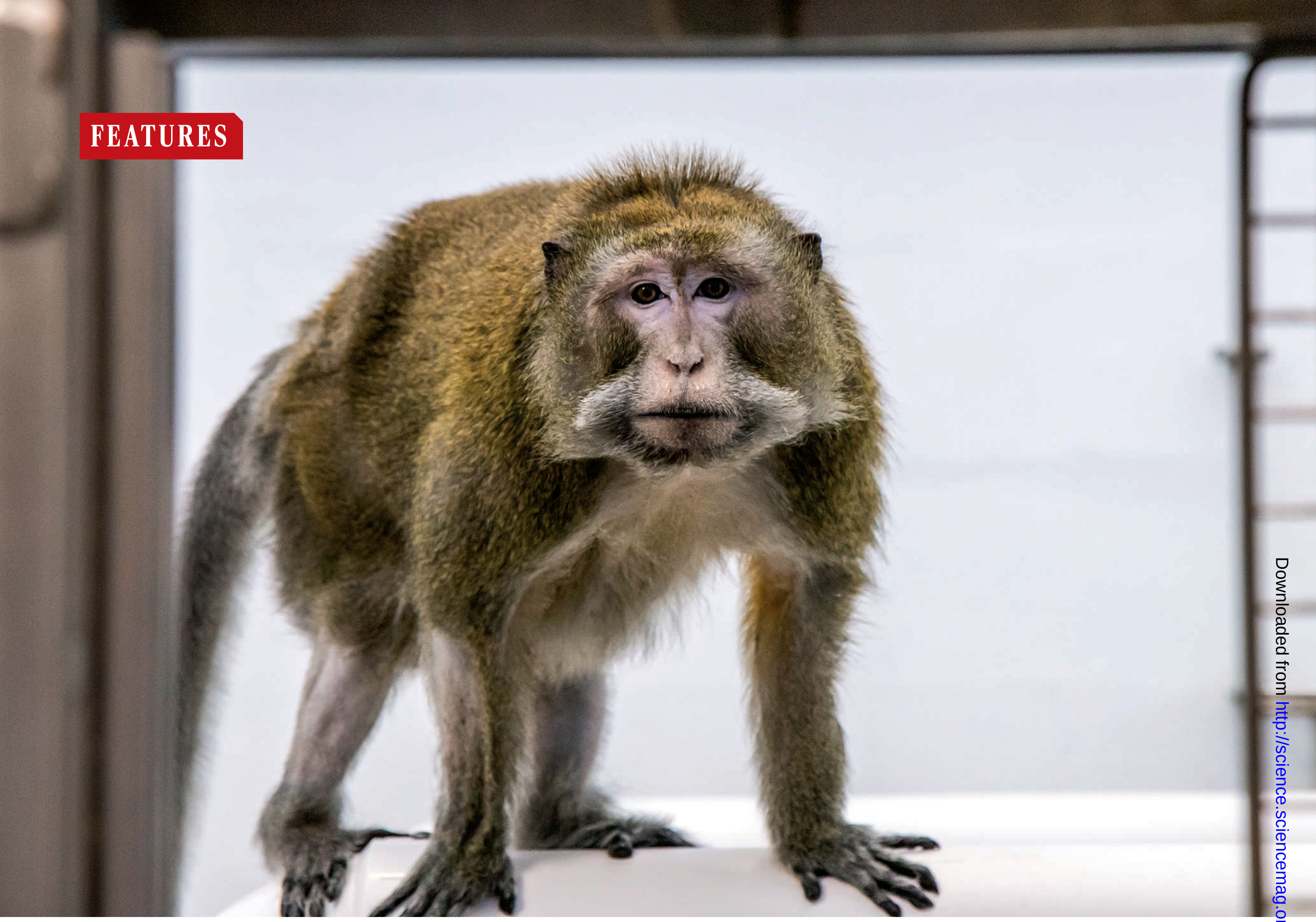
The Dutch high court implemented the ruling in May, halting all permit applications. It said the government needed to come up with a better system and a long-term plan to reduce nitrogen emissions. In September, a high-level commission suggested some short-term fixes, which the government has asked the high court to review. One idea is to lower the daytime speed limit from 130 to 100 kilometers per hour, which would reduce emissions enough to restart some home building. (The entire construction sector contributes just 0.6% of nitrogen emissions.) The government also wants to require changes in animal feed that would reduce nitrogen levels in manure and to buy out some farms near nature reserves. But the commission warned that deeper emissions cuts would require hard choices.

Some scientists and environmental groups say the Netherlands should move to circular agriculture: Farms should only produce as much manure as they can use to fertilize nearby fields; cows should graze rather than be fed nitrogen-rich, imported soy; and pigs and poultry should eat food waste. That would mean 50% fewer animals, says Natasja Oerlemans, head of agriculture for the World Wildlife Fund–Netherlands in Zeist. "We should use this crisis to transform agriculture," she says, adding that it will require several decades and billions of euros to reduce the number of animals.

LTO Netherlands in The Hague, which represents 35,000 farmers, endorses the concept of circular agriculture, but cautioned against "hasty measures." One new grassroots group, the Farmers Defence Force, contests the RIVM model's calculations of how much nitrogen from farms is deposited on nature reserves. RIVM has defended its model, which was peer reviewed before its 2015 launch. But it will ask an external committee to review both the model and the national nitrogen monitoring network.

Candel thinks EU courts might impose similar decisions on other European nations in the future. But for now, Dutch farmers will likely face tougher nitrogen restrictions than those in neighboring countries. That will rankle, especially because cross-border pollution is part of the problem, says Wim de Vries, who studies nitrogen impacts at Wageningen. About one-third of the nitrogen pollution deposited in the Netherlands comes from other countries, he says. "Nitrogen spreads everywhere." ■

FEATURES



READY TO RETIRE?

A nascent movement to send aging research monkeys to sanctuaries divides the biomedical community

By **David Grimm**, *in Winamac, Indiana*

It's been a long road to retirement for Bush the monkey—and not just because he's spent the past 15 hours in the back of a van motoring red-eye from New Jersey to Indiana.

For nearly his entire life, the 23-year-old macaque lived in a lab at Princeton University. There, researchers conducted MRI scans on him to understand which parts of the brain perceive faces, and he spent much of his time in an indoor cage. In 2017, with Bush suffering from arthritis and nearing the end of his life span, the lab decided to send him to a sanctuary.

"We had a very deep emotional relationship with Bush," says Sabine Kastner, a Princeton neuroscientist who oversaw studies on the monkey. "We were all very sad the day he left, but we were happy for him."

The university took more than 2 years to find a sanctuary it felt could provide high-quality care for Bush and had space for him. Finally, on 1 October, Princeton animal resource staff checked Bush's vitals, made him a care package of his favorite toys and treats, and placed him in a van that would bring him here to Peaceable Primate Sanctuary, a former farm amid sugar maples and cornfields where he'll spend the rest of his days.

Bush's big move is part of an unprecedented retirement collaboration: Princeton, along with Yale University, has just partnered with the sanctuary to ensure

PHOTO: DAVID KELLY CROW

Corrected 10 December 2019. See full text.

Published by AAAS

Bush, a cynomolgus macaque, was a research monkey at Princeton University for nearly 20 years.

both schools can seamlessly retire more monkeys there in the future. It's a sign of increasing interest in sending former research monkeys to sanctuaries instead of euthanizing them or transferring them to another project. A growing number of scientists say retirement is the right thing to do for these social, intelligent creatures, and it can be cheaper than keeping the animals in labs. "We want to do right by these animals," says Peter Smith, associate director of Yale's Animal Resources Center. "It's good for them, and it's good for the people who have spent their time caring for them."

Yet the effort faces many obstacles. More than 100,000 monkeys are in U.S. research facilities, and retiring even a fraction is a challenge. Labs often can't afford it or can't find a sanctuary they trust or that has space. And some primate researchers say sending monkeys to sanctuaries is simply a bad idea. Every one of these animals could contribute to crucial research, they argue, because monkeys can offer a deeper understanding of how our minds work as well as speed the search for cures for Ebola, Alzheimer's, and other diseases. Critics also fear that even talking about retirement could eventually lead to all monkeys disappearing from biomedical studies, as happened with chimpanzees. "I don't know of any monkeys that are not needed in biomedical research," says Amanda Dettmer, a comparative psychologist and primate researcher at Yale.

The discussion has grown even more heated in the past few months. Animal activist groups have pushed legislation that, if passed, would compel federal agencies to draft retirement plans for monkeys and other lab animals. "Taxpayers bought these animals, and we want the government to give them back," says Justin Goodman, vice president of the White Coat Waste Project, a Washington, D.C.–based group that has found an increasingly sympathetic ear in Congress by painting animal research as a misuse of tax dollars.

Meanwhile, according to a National Institutes of Health (NIH) report released in 2018, planned use of research monkeys was expected to continue to rise; the number used in experiments reached a record high in 2017 according to the United States Department of Agriculture (USDA), even though the total number held in U.S. labs has declined slightly over the past decade. Higher demand could cause a space crunch at biomedical facilities and expand the pool of older monkeys, making the question of retirement more urgent. All of this has left individual labs caught in the middle, struggling with whether to retire their monkeys—and the best way to do so.

As Bush, groggy and anxious, prepares to enter his new digs, it's unclear how many others will follow in his footsteps.

THE DIRT ROAD LEADING into Peaceable Primate opens up into 30 hectares of patchy grass bordered by tufts of woods. On the left, behind a chain-link fence, a couple of baboons chase each other through open concrete cylinders in an outdoor play yard. On the right, a single rhesus macaque with fluffy auburn fur scales a wooden climbing structure inside another play area, this one attached to a green, ranch style building with a red roof. Bush waits just inside.

Animal care staff at Princeton University prepare to transfer Bush to the van that will bring him to Peaceable Primate Sanctuary.

He's still in the wooden shipping crate he has been in since leaving Princeton. He pushes a couple of fingers through the wire mesh on top and peeks out cautiously, revealing a chestnut coat, stubby triangular ears, and bushy gray hair below his nose, giving the impression of a Mark Twain mustache. "I know, it's superscary," coos Laura Tarwater, head of animal care here. "Don't worry, we'll get you out of there soon."

Another Laura paces nearby: Laura Conour, director of lab animal resources at Princeton. She flew here to make sure Bush arrived safely, and she hovers around his crate like a nervous mother dropping her kid off at college for the first time. "He likes grapes—he likes to peel them," she tells Tarwater. "He likes the crunch of banana chips."

Tarwater and an animal caretaker shuffle Bush's crate farther inside the building. Here, 13 macaques live alone or paired in cinder block and wire runs the size of small bedrooms, filled with perches, swings, and rubber toys. Some animals are squat, long-tailed cynomolgus macaques (or "cynos") like Bush; others are lankier, short-tailed rhesus macaques. A few squawk and climb their enclosure to peek at the newcomer.

Tarwater and the caretaker set Bush's crate down firmly against the gate of his new run and open the doors on both to let him inside. He hesitates for a few minutes and then gingerly makes his way in. "It's his arthritis," Conour says. "He's probably stiff from the ride."

Like Bush, every animal in the building came from a research facility. Sixty-five percent of monkeys used in NIH-funded projects are rhesus; they're typically involved in brain studies or tests of therapies such as vaccines. Another 15% are cynos, which constitute most monkeys in industry labs that test drug safety. Thirteen other primate species—including baboons and marmosets—make up the rest (see graphic, p. 1184).

None has been part of the retirement conversation until recently. For the past decade, the focus has been on chimpanzees. In 2010, in response to public and congressional pressure, NIH commissioned a report that concluded most biomedical research

PHOTO: DAVID KELLY CROW

Corrected 10 December 2019. See full text.

on chimps was unnecessary. Five years later, the agency announced it would no longer support invasive research on the animals and would retire all its chimpanzees (*Science*, 16 June 2017, p. 1114).

Animal activist groups are following the chimpanzee blueprint in their public relations campaigns, pushing for monkey retirement as a way to reduce—or one day even eliminate—use of the animals in research. Last year, White Coat Waste orchestrated a congressional letter that urged federal labs to disclose what they did with monkeys after experiments ended. In May, the group got members of the House of Representatives to include language in a report attached to a proposed NIH spending bill that asks the agency to reduce its use of the monkeys it owns and formulate a plan to retire them. That same month, House members working with the group introduced a bill that would require all federal agencies—which own about 9000 monkeys—to create policies to retire lab animals. (The Senate followed with its own bill.)

None of the proposed legislation has passed, and none would force agencies to retire their monkeys. But researchers have been blindsided before, when efforts by White Coat Waste ended studies on squirrel monkeys, cats, and dogs at several federal facilities. Adding fuel to the retirement conversation, a January memo to Congress from the Department of Health and Human Services—which includes NIH, the Centers for Disease Control and Prevention, and the Food and Drug Administration—stated that it is working toward retiring lab animals. "We are supportive of the adoption of cats, dogs, and primates when relocation is safe and medically appropriate," the agencies wrote.

The proposals target only federal research, so they wouldn't affect academic institutions. But a few are moving ahead with their own plans.

A HALF-HOUR HAS PASSED, and Bush seems more at ease in his new quarters. The sanctuary staff open a small door that leads outside to his own play area, with barrels to crawl into, swinging ropes made of firehoses, and bright orange balls stuffed with treats. Bush inches toward the opening and stops. He has never been outdoors. He has only seen the sun through a skylight. "Come on, Bushy!" Conour calls. But he just stares through the door at the world beyond.

"All of them eventually go outside, but we don't force them," says Scott Kubisch, the sanctuary's founder and director. "Being retired is all about having choices."

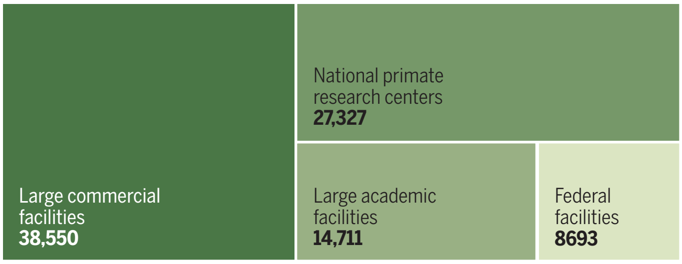
Kubisch talks about the sanctuary with the affection of someone who has spent more than 2 decades building it from nothing, largely with his own hands. He had the idea for Peaceable Primate in 1996, when he was a primate keeper at the Lincoln Park Zoo in Chicago, Illinois. He wanted to do something for laboratory monkeys after their research days ended. So he started a nonprofit, asked friends to collect coins in monkey-shaped piggy banks, and took money from his retirement savings to buy land here, more than 2 hours south of Chicago, near a farm where he lived as a kid.

"On weekends, I would come down and build the chicken coop and put in fences. Every tree you see here, I planted," he says. "People at the zoo thought I was crazy."

In 2014, a friend left Kubisch a large donation in her will. With it, he created an endowment for the sanctuary and constructed most of the remaining buildings. He originally intended Peaceable Primate as a baboon refuge, and the sanctuary took in its first three baboons in 2016. "But we were getting a lot more calls from universities trying to retire macaques," he says, so the sanctuary began to take them in 2018. It now has 18 baboons and 14 macaques, including Bush.

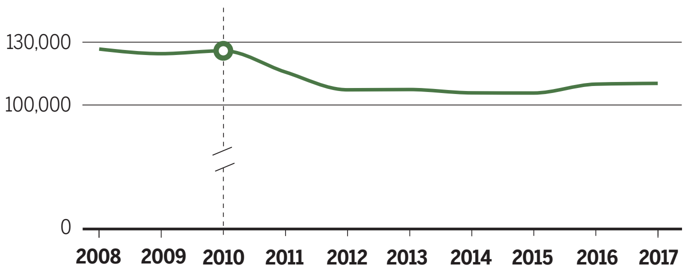
Where the monkeys are

Most U.S. research monkeys live in four main types of facilities. In 2015, the last year relevant data were available, industry labs held the most.



Primed to rise

The total number of monkeys held in U.S. facilities has trended slightly downward over the past decade, but NIH expects these numbers to rise as researchers request more monkeys for studies on topics such as vaccines and the aging brain.



Such calls represented a shift: Many U.S. biomedical researchers have traditionally viewed sanctuaries suspiciously, fearing they are run by animal rights activists who will tar labs in the press to gin up public sympathy and donations. Other scientists have worried about the quality of care, noting that unlike labs, sanctuaries don't have to register with the USDA, which requires regular inspections. (Many sanctuaries do seek membership with the North American Primate Sanctuary Alliance, which imposes strict standards of care.) When you send an animal to a sanctuary, says Cindy Buckmaster, chair of Americans for Medical Progress and former director of one of the country's largest animal care and use facilities at Baylor College of Medicine in Houston, Texas, "you're putting them in a situation full of question marks."

Kubisch has addressed those concerns head-on. He is careful to neither extol nor denigrate monkey research. "Our stance is that we don't take a stance," he says. He also has registered Peaceable Primate with USDA and always has at least three animal care staff on the premises.

Kubisch has tried to keep costs down as well. Sanctuaries can charge up to $25,000 to care for a monkey, depending on how many years it has left. (Macaques live an average of 27 years.) Peaceable Primate asks for about $2500 per year per animal. That's much more expensive than euthanizing a monkey, but it's cheaper than keeping it at a university—"a fifth of what Yale pays," says Smith, who oversees about 150 monkeys there.

Both he and Conour say monkey retirement had always been informal and piecemeal at their universities. Each has retired only about a half-dozen monkeys during their tenures. When they learned about Peaceable Primate, they persuaded their institutions to make a large financial investment. (Neither school would say how much.) Conour and Smith say their universities felt that forging a relationship with a single, well-regarded sanctuary would smooth the path of retirement.

The result: a newly cleared plot of land just west of the macaque house. It's not

CREDITS: (GRAPHIC) X. LIU/*SCIENCE*; (DATA) NATIONAL INSTITUTES OF HEALTH; U.S. DEPARTMENT OF AGRICULTURE

From his indoor run, Bush peeks out at his play area at Peaceable Primate Sanctuary.

much to look at today—a few cinder block bricks and a whole bunch of dirt. But by the end of the year, it will be a new macaque building, with the same large runs and outdoor play area as the first one. "In their indoor runs alone, they're going to have 10 to 20 times as much space as they do at Yale," Smith says.

Bush will eventually move in here, and the house will be reserved for animals from Yale and Princeton. "It's going to be the Ivy League building," Smith jokes. "They'll all be wearing tweed coats."

The universities' arrangement covers the cost of the building and lifetime care for six macaques from each school, even as old ones die and new ones come in. "We now have a pipeline to retirement," Smith says.

But not everyone in the biomedical community is likely to get on board.

"IMAGINE YOU'RE A 70-YEAR-OLD human who knows everyone in your neighborhood, and then people pack you in a van and take you to a strange, new place where you don't know anyone," says Dettmer, the Yale primate researcher. "Even if it's beautiful, being ripped away from everything you know can be devastating." She points to an incident about 5 years ago, when 13 elderly research chimpanzees were transferred to a federal sanctuary in Louisiana. Within 2 years, nine had died. The sanctuary said the chimps were sick and elderly, but many people in the biomedical community blamed the stress of relocation.

Dettmer says the decision to retire chimpanzees to sanctuaries was a mistake, one she fears is being repeated with monkeys. Even older animals can be used in studies of the aging brain and body, she says. "There's no such thing as a surplus monkey."

She, Buckmaster, and others also say none of the pending retirement legislation is realistic—it offers no money for retirement and creates no federal sanctuary, as was done with chimps. To retire and care for just the few thousand monkeys in federal facilities would cost more than $400 million, according to Speaking of Research, an international organization that supports the use of animals in labs. And because only about a dozen U.S. sanctuaries now take monkeys, the sanctuary system doesn't have space for even a fraction of them.

Critics like Dettmer (who is active in Speaking of Research) also worry about a slippery slope. Talk of chimpanzee retirement went hand in hand with removing those animals from biomedical research—a concern "in the forefront of researchers' minds," regarding monkeys, Dettmer says.

Dettmer says she understands the emotional appeal of monkey retirement but argues that the needs of humans should come first. "We're not just concerned about the welfare of animals," she says. "We're concerned about the welfare of society."

IN THE SWIRL OF OPINIONS, labs are trying to find their own way forward. Some universities have already reached out to Smith to see whether they can create a retirement pipeline, too. Conour says a drug company has contacted her. "Right now, we're the cool kids on the block," she laughs.

Johns Hopkins University in Baltimore, Maryland, is taking a different route. It's considering creating its own sanctuary, on a farm in the Maryland countryside. Having its own facility would be cheaper and easier than sending monkeys to an outside sanctuary. What's more, university staff "would see the animals they cared for and came to know, sometimes since they were babies, enjoying their postresearch lives," says Eric Hutchinson, associate director of research animal resources at the university, who came up with the idea. "And the public would see that an institution can be committed to both animal welfare and high-quality biomedical research."

For labs going it alone, the Research Animal Retirement Foundation may be able to help. Founded by former monkey lab manager Rachele McAndrew in Gilbert, Arizona, the organization is trying to raise funds for scientists who want to retire monkeys. It also offers advice on finding and working with sanctuaries. "We want to be a one-stop shop for labs interested in retirement," McAndrew says. Smith says the interest is there. "I think it's going to be a fairly hot topic for the foreseeable future."

Even Buckmaster, with her concerns about the sanctuary community, has found one she trusts and has retired close to a dozen monkeys there. "My pipe dream is for a small network of government-owned sanctuaries spread throughout the country that could take any type of research animal," she says. "The public wants that. We want that. These animals deserve that."

Kubisch has his own dream, hoping to continue to grow his sanctuary. "I think we could eventually house up to 500 animals," he says. "I want to be the go-to place for retirement."

Tomorrow, Bush will venture outdoors for the first time. He'll wince at a blast of wind against his fur. He'll stare curiously at birds as they chirp overhead. And he'll push his hand into the dirt, feeling it envelop his fingers. Then he'll look behind him at the small door to his run, perhaps pondering whether he should go back inside—or remain in the sun. ■

PHOTO: ALYSSA SCHUKAR

INSIGHTS

BOOKS *et al.*

AAAS/SUBARU SB&F PRIZES FOR EXCELLENCE IN SCIENCE BOOKS

Wishlist-worthy books for young readers

From cryptic ciphers and wall-climbing robots to dinosaur digs and visits to the planetarium, this year's finalists for the Science Books and Films (SB&F) Prizes for Excellence in Science Books are packed with fun facts, easy-to-do experiments, and plenty of creepy-crawly creatures. Sponsored by Subaru and the American Association for the Advancement of Science (AAAS, the publisher of *Science*), the SB&F competition celebrates high-quality science books for young readers. Read on for reviews of the finalists, written by the staff of the *Science* family of journals, with help from a few friends. —**Valerie Thompson**

CHILDREN'S SCIENCE PICTURE BOOK

Follow That Bee!

Reviewed by **Anand Balasubramani**[1]

Follow That Bee! introduces kids to bees and urban beekeeping. The book describes an afternoon that five friends—Martin, Nick, Yulee, Sally, and Pedro—spend with Mr. Cardinal, an urban beekeeper. Mr. Cardinal starts by telling the kids that bees like to eat both nectar and pollen from flowers and explains how to set up a bee-friendly garden. The kids learn fun facts about bees and the importance of bees as pollinators. (My favorite factoid was that bees cannot see the color red and thus prefer purple, blue, and yellow flowers.)

Mr. Cardinal then introduces the kids to the roles of different bees in a hive and how they talk to each other. When they learn that bees communicate with each other by dancing, the children start dancing too.

Later on, a bee stings Nick's foot. Mr. Cardinal, pulling out the stinger and cleaning the wound, uses this as an opportunity to remind the kids that they should be very careful around bees. Toward the end of the visit, he shows the kids how he uses smoke to calm the bees when he collects honey from their hives, and they help him bottle the honey.

Scot Ritchie's book contains some superb illustrations of bees and the hives they live in. It also does an excellent job of explaining the importance of urban beekeeping and offers tips for how both kids and adults can give bees a helping hand, such as by limiting the use of pesticides and by setting up baths for bees to drink water from in our gardens.

Follow That Bee! A First Book of Bees in the City, Scot Ritchie, Kids Can Press, 2019, 32 pp.

When Sue Found Sue

Reviewed by **Jennifer Sills**[2]

For children who constantly fill their pockets with treasures, there is the story of Sue Hendrickson. A child who loved to find things, Hendrickson grew up to discover a skeleton of a *Tyrannosaurus rex*.

In *When Sue Found Sue*, we learn that Hendrickson, who was shy and isolated in her early years, had little formal schooling but lots of curiosity. The story briefly touches on Hendrickson's early career diving for lost ships and mining for amber before turning to her work digging for dinosaur fossils in South Dakota. There, she and a team of paleontologists—all driven by a passion for finding things—faced intense heat and spartan accommodations.

One day, Hendrickson ventured off alone and spotted bones she knew belonged to

ILLUSTRATION: GABY D'ALESSANDRO

a *T. rex*. The team named the *T. rex* “Sue” in her honor. It is now displayed in a museum Hendrickson frequented as a child.

Quotes from Hendrickson give the tale an added boost of realism, but the text, in staying true to Hendrickson’s life, might occasionally leave young readers confused. (References to Dominican amber mines and “a long dispute about ownership” of the *T. rex*, for example, go unexplained.) Nonetheless, children will relate to Hendrickson’s joy in discoveries big and small, highlighted by splashes of yellow in the vibrant watercolor illustrations.

Hendrickson’s transition from a solitary outsider to an integral part of a team gives the story another note of optimism. One way to read the book’s title is as a straightforward description of the climactic moment when Sue the scientist found Sue the *T. rex*. But perhaps as Sue raced back to share her findings with the group of like-minded treasure hunters, she also found herself.

When Sue Found Sue: Sue Hendrickson Discovers Her T. Rex, Toni Buzzeo, Illustrated by Diana Sudyka, Abrams, 2019, 32 pp.

Moth

Reviewed by **Brad Wible**[3]

Moth is a storybook illustration of a textbook tale of environmental degradation, then rehabilitation, and the natural selection driven by both. The tale is vividly brought to life in spellbinding watercolor collage renderings of swirling *Geometridae* and landscapes caked in soot.

We learn that peppered moths (*Biston betularia*) in the United Kingdom had a liking for trees covered in lichen, their light-and-dark speckled wing and body coloring camouflaging them as they rested. The more charcoal-colored among them were, as any hungry bird would have told you, more conspicuous and thus extremely rare.

But the Industrial Revolution turned the world upside down, literally turning white to black. Coal-fired factories drowned the sky and everything under it with soot. Trees blackened, lichen succumbed to sulfur dioxide, and the charcoal-colored moths became less conspicuous than their lighter relatives. Again, the birds noticed, and the once-prominent speckled moths became far less common than their charcoal-colored kin.

Eventually, we began to recognize the myriad ills of pollution. The air got cleaner. The trees and lichen recovered. And once again the lighter-speckled moths became more successful at blending in and survived.

Evolution is not always so tidy. But together, Thomas’s words and Egnéus’s illustrations introduce the contours and landmarks of this story in an elegant and engaging fashion. *Moth* also conveys broader lessons to guide young readers on a journey into a life in science, or just a life well lived: “Scramble through a forest… Be silent. Be still. Look closely…and hope.”

Moth: An Evolution Story, Isabel Thomas, Illustrated by Daniel Egnéus, Bloomsbury, 2019, 48 pp.

Butterflies in Room 6

Reviewed by **Caroline Ash**[4]

Metamorphosis—of ugly ducklings into swans, of jellylike spawn into frogs, of caterpillars into butterflies—always seems miraculous. In this book on insect metamorphosis, Caroline Arnold tells the story of Mrs. Best, a kindergarten teacher who brings a tiny vial of butterfly eggs into her classroom. Her students supply a vivarium with special caterpillar food so they can watch the metamorphosis of the eggs into caterpillars, then pupae, and finally glorious adult painted ladies. The book takes the reader through the course of the children’s project, with a series of fine photographs showing the details of each stage in the life cycle of the butterflies. The exciting anticipation of each transformation is summarized in carefully considered text and culminates, of course, with the day the exquisite adults emerge from the pupal case, unfurl, and stiffen their patterned wings. Beautiful close-up images let the readers examine details of the insects’ anatomy and learn about butterfly biology.

Finally, a warm day arrives, and it is time to release the butterflies. The dazed insects first walk onto the children’s hands before lifting off to disappear over the horizon. Fortunately, some hang around to appreciate the school garden’s flowers.

It would have been good for *Butterflies in Room 6* to say a little more about why insects are having such a tough time now, as well as more about their role in pollination and human food security. Still, it is an excellent book, sure to generate discussion and flights of imagination among humans who are similarly poised for big changes.

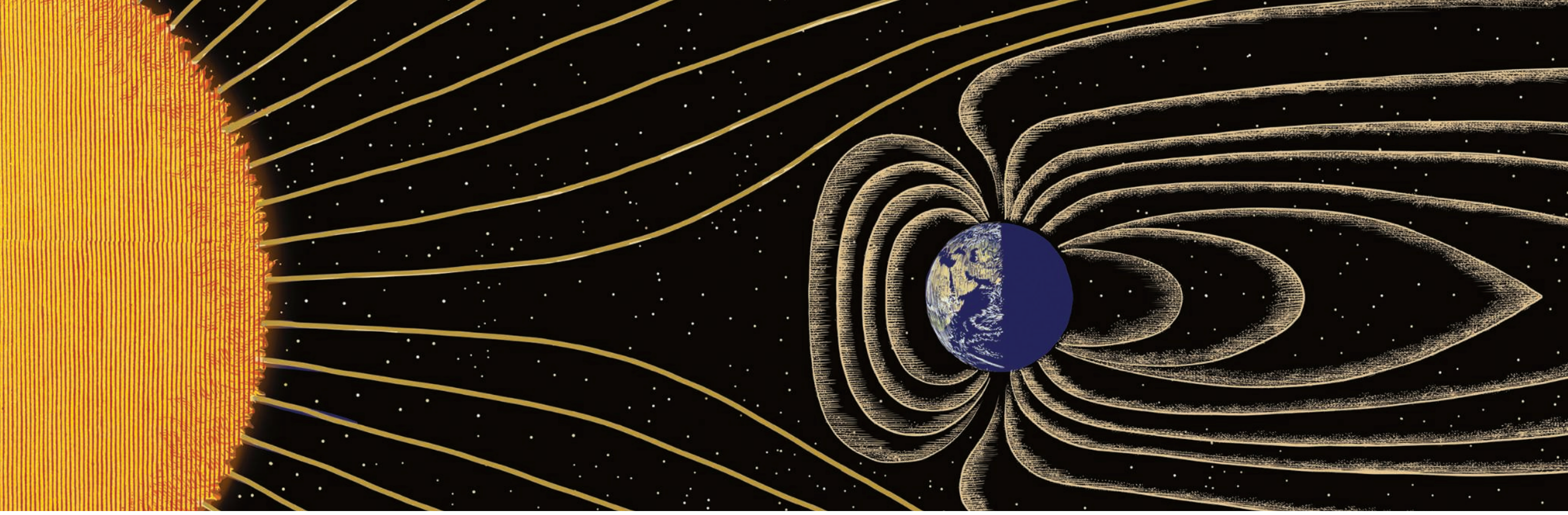
Butterflies in Room 6: See How They Grow, Caroline Arnold, Charlesbridge, 2019, 40 pp.

MIDDLE GRADES SCIENCE BOOK

Planetarium

Reviewed by **Hilary Stewart**[5]

Planetarium is designed to replicate the experience of walking through a museum exhibition. It succeeds in this goal, capturing the residual wonder that one feels when stepping out of a quiet planetarium



Particles from the Sun compress Earth's magnetosphere on the side facing the Sun and stretch it into a long tail on the side facing away from the Sun.

into the bright light of day. Raman Prinja takes readers on a tour of space, starting from our own Solar System. We stop at the Sun and every planet along the way, then travel through asteroids, comets, and dwarf planets, before moving on to exoplanets, other stars, our Galaxy, and finally our Universe. Each page is accompanied by Chris Wormell's beautiful illustrations, which make the book feel truly timeless.

Like any good museum display, each page is bite-sized and accessible to its target age group (8- to 12-year-olds). Prinja makes liberal and clever use of analogies to convey just how big (or small, or difficult) things are. He answers questions before the reader can even ask them, often with a nice touch of humor: How do we find other planets? What would it be like to stand on Venus? Who decided on the constellations we use today? What holds the arms of a spiral galaxy together?

The book concludes with theories of how our Universe will end, which range from "The Big Rip," where the Universe's growth accelerates until it tears itself apart; to "The Big Crunch," a contraction of the Universe down to nearly nothing; to "The Big Chill," the continual expansion of the Universe until it slowly cools and dies. This should feel more grim than it does, but Prinja cuts the tension with wonder at the mysteries that are left to uncover and a persistent sense that we are on the cusp of knowing more.

Planetarium, Raman Prinja, Illustrated by Chris Wormell, Big Picture Press, 2019, 104 pp.

Eye Spy

Reviewed by **Pamela J. Hines**[6]

The table in my hallway is blue. My son swears it's green. My nephew says it's brown. Vision helps us interpret the world around us, but what if we each see the same thing differently? *Eye Spy*, a child's lift-the-flap book, gives readers a glimpse into the lives of animals whose views of the world are even more divergent than my family's.

The woodcock has eyes on the sides of its head that deliver a full-circle view. The chameleon, disorienting to look at, has an equally disorienting view of its own world, with eyes that can work together or roll around independently, as needed. The book is filled with fun facts and relates field of view, color perception, and sharpness of focus to an animal's place in the pecking order: predator, prey, vegetarian, frugivore. The lift-the-flap approach brings an element of surprise for a child looking through the eyes of the animals drawn on the page, although the detailed factual summaries inside the flaps might require a bit of adult interpretation for the younger child.

In the end, I was reassured to learn that my perpetual nearsightedness, although perhaps not normal for humans, is how most chimpanzees see the world. (Maybe that's why I like fruit so much?)

Eye Spy: Wild Ways Animals See the World, Guillaume Duprat, What on Earth Books, 2018, 36 pp.

Kid Scientists

Reviewed by **Hadassah Nusinovich Ucko**[7] *and* **Yevgeniya Nusinovich**[8]

Kid Scientists by David Stabler is a biography about the childhoods of famous scientists. The featured scientists—including Jane Goodall, Albert Einstein, and Stephen Hawking—are presented at around the same age as the target audience, making it easier for readers to identify with them. The book's main goal, which is admirably achieved, appears to be supporting young people's interest in science.

Each story starts with that particular scientist's early inspiration and is accompanied by simple activities that any child can do, such as stargazing. An essay on Katherine Johnson, for example, describes her predilection for counting household items, such as dishes, long before she became a professional mathematician.

Stabler carefully avoids favoring any one group of people, achieving an even balance between scientists of different races and genders, which makes it easier for children of various backgrounds to find role models who look like them. Although the author's efforts to emphasize the stories of women and people of color appear obvious to an adult reader, children are unlikely to notice this as a deliberate choice on the author's part. (Children may likewise fail to notice when the book glosses over important details, such as the influence of the World Wars on the scientific enterprise.) Occasionally, the book's attempts to communicate the barriers faced by women and minorities in science lack subtlety and can come across as heavy-handed (e.g., "because of her gender, [Rosalind Franklin] was never given proper credit for her discoveries"), but, overall, the stories read well.

In the end, *Kid Scientists* is an interesting collection of anecdotes, with stories about Benjamin Franklin's swimming fins, Ada Lovelace's obsession with flying horses, and other fun facts that most readers are unlikely to have encountered before. As such, most children should enjoy reading it, without noticing how much they are learning, as they internalize the idea that they, too, can grow up to be scientists.

Kid Scientists: True Tales of Childhood from Science Superstars, David Stabler, Illustrated by Anoosha Syed, Quirk Books, 2018, 207 pp.

Owling

Reviewed by **Laura M. Zahn**[9]

Birds are some of the most accessible wild animals. Observing them requires no special equipment. Although we may hear owls outside as we prepare to slumber, however, few of us see them regularly.

Focusing on the biology of these predatory

birds, this book is broken into four chapters that are accessible to most ages, although younger children will probably want to focus on the many fascinating owl pictures. Fun facts are tucked among the pages, often with photographic illustrations. Did you know, for example, that owls have feathered eyelids?

Author Mark Wilson is a wildlife photojournalist who also cares for owls that cannot live in the wild. This access lends itself to stunning images of the species discussed in the book, which include photographs of their habitat and identifying marks, as well as many images of nestlings.

The book discusses some avian commonalities, as well as how owls differ from other birds. One section specifically introduces the 19 owl species of North America, including how and where to find and identify them. Noting that owls are probably closer than one might think and that many are active during the day, one of the highlights of the book is the discussion of how to study one's surroundings to identify the likely presence and location of a nearby owl. (Hint: Look for their excrement and remains of their meals, known as owl pellets.)

Owling presents several case studies of people who work with owls, providing role models and offering potential future careers for its younger readers. Anyone interested in identifying and learning about the owls with which we share the world will likely enjoy paging through this book, all the more so if it is used as an owl guide.

Owling: Enter the World of the Mysterious Birds of the Night, Mark Wilson, Storey Publishing, 2019, 122 pp.

HANDS-ON SCIENCE BOOK

Plantology

Reviewed by **Michael A. Funk**[10]

Go outside and look around. Unless your surroundings are buried in snow, you will likely see green leaves, budding flowers, and tough weeds. But have you ever stopped to compare leaf shapes, count the number of petals, or see if a taproot is hidden underground? Such activities are just the tip of the iceberg (lettuce!) when it comes to the experiments and observations suggested in *Plantology*.

Starting with the simple question "what is a plant?" this book delves into the vast and often surprising diversity of the kingdom Plantae. Each chapter covers an aspect of plant structure and function, laying out general principles but also offering surprising examples of when things get weird. Did you know that some orchids trap bees inside their flowers in order to glue pollen to them? Or that some plants do not have leaves at all and instead extract energy from fungi through their roots? These facts collectively emphasize that every plant is worth a careful look, as millions of years of adaptation have left nearly every evolutionary route explored and few rules unbroken.

Aside from the information conveyed in the diagrams and (quite complex) descriptions and anecdotes, this book offers a simple imperative for budding naturalists: Go outside and document everything you see. Colored pencils, a nature journal, and a hand lens would make great accompaniments for this book, as they are needed for many of the activities. Most of the exercises require little more than this equipment, time, and attention, although there are a few ideas for those with access to a garden plot and a kitchen. The final chapter offers a brief introduction to how humans use plants as crops and, if young readers are hooked on plants by then, some career advice.

Plantology: 30 Activities and Observations for Exploring the World of Plants, Michael Elsohn Ross, Chicago Review Press, 2019, 128 pp.

Chameleons see colors better than we do, helping them catch prey and create convincing camouflage.

Science in a Jar

Reviewed by **Seth Scanlon**[11]

Do you have a surplus of glass jars clogging your cupboards and children who are curious and excited about the world? If so, Julia Garstecki's collection of more than 35 household experiments may be a great addition to your library. This book is divided into chapters on life science, chemistry, earth science, physical science, and environmental science. Similar to Jonathan Adolph's 2018 book *Mason Jar Science*, the premise is that all of these experiments are to be conducted using glass jars of varying sizes. Although the aim is to use common household materials (stock up on food coloring—you'll need it!), experiments occasionally call for items that are unlikely to be lying around (e.g., iron filings and a dual-probe thermometer).

For each experiment, the central concepts being tested are artfully conveyed using bright, inviting photos and accessibly written introductions, followed by a materials list and clear step-by-step instructions. Each experiment closes with an "Observations" section, which provides explanations and further questions. Fun facts on tangential topics—from earthworm poop to the autonomous sensory meridian response—make an occasional appearance.

Many classic household experiments—such as dissolving eggs in acid, synthesizing slime, and generating explosions of vinegar and baking soda—can be found among the book's 120+ pages. Although most of the experiments listed are meant to be conducted indoors, several encourage children to go outside and explore the natural world. Where *Science in a Jar* really shines is with its simplest experiments, which manage to explain broadly interesting concepts—such as why puppies snuggle when they sleep and how clouds form—in a fun and accessible way.

Science in a Jar: 35+ Experiments in Biology, Chemistry, Weather, the Environment, and More!, Julia Garstecki, Quarto, 2019, 128 pp.

Can You Crack the Code?

Reviewed by **Adrian Cho**[12]

Readers who think it would be fun to send secret messages—i.e., practically all of them—will enjoy this book, which is both a brief history of encryption and a beginner's how-to manual. Ella Schwartz, a cybersecurity expert, uses a breezy style that itself is never cryptic, and Lily Williams's illustrations make the concepts readily accessible.

The book begins with a simple substitution cipher, used by Julius Caesar, in which every letter of the alphabet is signified by another letter a specific number of spots away. It then works through more-complicated substitution ciphers, symbolic ciphers, and ciphers in which a message is spelled out in the first letters of specific words in a specific book. Schwartz poses puzzles so the reader can try the schemes. She describes Enigma, the complex cipher used by the Germans, and cracked by the British, during World War II, and the basic idea behind internet encryption.

Along the way, Schwartz provides fascinating examples of codes that have long defied

ILLUSTRATION: GUILLAUME DUPRAT

Published by AAAS

decryption. For example, in 1990 the U.S. Central Intelligence Agency installed at its headquarters in Langley, Virginia, a sculpture in which artist Jim Sanborn encrypted a message. Even the agency's cryptographers have yet to completely decipher it.

The book might have gone a tad deeper and explained how digital messages consist of binary numbers and how one can be encrypted, for example, using a string of random zeros and ones that only the sender and the receiver share. Such a discussion might have set up the chapter on internet encryption more concretely.

But that's a quibble. The book should inspire students who enjoy puzzles to invent their own ciphers. As Caesar might have put it, vjku dqqm ku c nqv qh hwp!

Can You Crack the Code? A Fascinating History of Ciphers and Cryptography, Ella Schwartz, Illustrated by Lily Williams, Bloomsbury, 2019, 128 pp.

George Washington Carver for Kids

Reviewed by **Monica Bradford**[13]

Most people remember George Washington Carver as a Black inventor who created close to 300 peanut products. Peggy Thomas's book moves beyond that one-dimensional view as she guides the reader through Carver's life story. Born in the early 1860s to enslaved parents, Carver overcame poverty and segregation to devote his life to improving the livelihood of impoverished Black farmers. Combining his early love of nature and painting, Carver became a naturalist, an ecologist, and a conservationist long before these were valued disciplines.

Sprinkled through the book are 21 activities related to Carter's life experiences and the challenges he faced. Some of these easy-to-follow activities encourage readers to act like a naturalist by, for example, making their own herbarium. Others allow readers to experience tasks required for daily survival: turning a gourd into a bowl, cooking with weeds, and making blocks like those Carver used to construct a sod home.

As the first Black student to attend Iowa Agricultural College (IAC; now Iowa State University), Carver faced segregation and isolation until an act of solidarity helped break the ice with the other students and faculty. The activity associated with this event encourages readers to form a welcoming committee for new students. Another activity focuses on learning to deliver a speech, a skill that Carver developed at IAC and later used to teach farmers, politicians, and leaders of industry about plants, soils, and the potential of natural products.

Clearly, the knowledge Carver shared in the early 1900s still resonates today. Hopefully, young readers will be inspired by his resilience, thirst for learning, and passion for improving lives.

George Washington Carver for Kids: His Life and Discoveries with 21 Activities, Peggy Thomas, Chicago Review Press, 2019, 136 pp.

YOUNG ADULT SCIENCE BOOK

How to Walk on Water and Climb up Walls

Reviewed by **Marc S. Lavine**[14]

At the intersection of fluid, solid, and gaseous flow; animal motion and mechanics; applied mathematics; and robotics lies the research of David Hu. Through stories about researchers he has interacted with over the course of his career and a biographical view of his own projects, Hu shows readers that we still have a lot to learn about animal locomotion.

As humans, we are capable of traversing a range of terrain, including mud, sand, and rocky ground. We can move through water or air. There are many animals, however, that excel in environments where we struggle. The indestructible cockroach, for example, not only can survive being squashed to a quarter of its original height, it will alter the shape of its legs so that it can continue to move forward. It also has antennae that it can move at full speed even when surrounded by obstacles. These abilities are reasons why the cockroach is used as a model organism for the development of inexpensive robots that might one day be used in search-and-rescue missions in hazardous environments.

The combination of fluid mechanics and animal physiology helps explain the optimal length of eyelashes for filtering out debris, flying snakes that glide efficiently with no wing or web structure, and how ants form robust collective rafts. However, the study of motion goes beyond the movement itself. Animals such as lamprey reveal how motion can be controlled, not by a brain but by a series of central pattern generators that are sets of small interacting periodic motions. This allows a salamander to change from a walking motion to a swimming one just by altering the speed of the lead oscillations.

Studying animal motion in fine detail may seem wasteful to many, but as Hu convincingly shows, a detailed understanding of the interaction between biology and fluid mechanics is interesting in its own right and a rich source of design and engineering ideas.

How to Walk on Water and Climb up Walls: Animal Movement and the Robots of the Future, David L. Hu, Princeton University Press, 2018, 238 pp.

Never Home Alone

Reviewed by **Andrew Sugden**[15]

Although you may encounter reminders of life that flourishes outdoors—the occasional spider in the bath, the mold that sometimes sprouts on food—by and large, the only living beings you consciously encounter in your home are the other people in your life and your pets. After reading this book, you will probably see things quite differently.

Our homes are teeming with life. A microbial analysis of household dust samples from just 40 homes in North Carolina yielded nearly 8000 different microbial taxa, for example. Household microbial communities vary by habitat: As Rob Dunn writes, "…samples from pillowcases and toilet seats are different from each other, but perhaps not as different as you might hope."

Larger creatures are often hiding in plain sight as well. Dunn describes surveys of the arthropod communities in homes around the world. His own house was home to at least 100 arthropod species—mainly flies and spiders—and worldwide, the numbers of species stretch into the thousands.

Dunn considers how human history (and prehistory) may have shaped the microbial communities that have evolved to share our lives. He delves into research that enumerates the health benefits to humans of closer contact with natural biodiversity and, conversely, the problems that arise through the evolution of pathogens resistant to control measures. Perhaps above all, this book is a vehicle for conveying the story of how science is done—the quirks and collaborations that lead from one discovery to another, and the role of citizen science in advancing knowledge.

Never Home Alone: From Microbes to Millipedes, Camel Crickets, and Honeybees, the Natural History of Where We Live, Rob Dunn, Basic Books, 2018, 330 pp.

Superheavy

Reviewed by **Nathaniel Smith**[16]

Kit Chapman begins his story of the early hunt for so-called "superheavy" elements with Enrico Fermi, the scrappy Italian physicist who claimed, erroneously, to have discovered elements 93 and 94. As the first scientist to develop a technique for a phenomenon

called neutron capture, Fermi paved the way for a number of element discoveries to come. From Rome, we travel to Berlin for the discovery of nuclear fission in 1938 and then on to Berkeley, California, where scientists began the search for the transuranic elements (elements heavier than uranium) in 1939.

Chapman does an admirable job of bringing to the forefront the incredible contributions of women scientists to this endeavor. One such story is that of nuclear chemist Darleane Hoffman, who was told repeatedly that she should set her sights on becoming a chemistry teacher. Instead, she would go on to hold appointments at Oak Ridge National Laboratory, Los Alamos, and the University of California, Berkeley, and earn her place in history by confirming the existence of seaborgium (element 106).

In the final third of the book, Chapman takes readers on a tour of the modern world of element discovery. Tales from RIKEN in Japan, GSI in Germany, and other international laboratories complement stories from the early post–Cold War period, including anecdotes about collaborations between researchers at Berkeley and those in Dubna, Russia.

Although some transuranic elements have proven useful in medical science and other applications, the elements beyond 118 are very unstable. Practical use of these materials, then, is not the point. Even the honor of choosing the names of newly discovered elements—a topic into which Chapman dives perhaps too deeply—is not the point. It is the acquisition of knowledge, the drive to expand our understanding of the world, that many would say is the purpose of such an endeavor.

Superheavy: Making and Breaking the Periodic Table, Kit Chapman, Bloomsbury Sigma, 2019, 304 pp.

For a full-length review of *Superheavy*, see "Our autumn reading list," *Science* **365**, 972 (2019).

A plane flies over southeast Greenland as part of a NASA mission to monitor polar ice.

The Ice at the End of the World

Reviewed by **William E. Glassley**[17]

Although Greenland occasionally figures into news stories and debates about rising sea level and a warming Earth, the relevant historical background rarely enters into such discussions. Consequently, many fascinating elements of the island's story that could capture the interest and attention of the public have lain dormant. Jon Gertner's compelling book, *The Ice at the End of the World*, addresses this paucity with intelligence and insight.

The book is divided into two parts. Part 1, "Explorations," describes expeditions that took place near the turn of the 20th century. Gertner conveys the psychological and physical struggles these individuals suffered through, as well as their exhilarating successes, with a realism that acknowledges the emotional and physical toll of these endeavors. In Part 2, "Investigations," Gertner chronicles the evolution of scientific studies of the Greenland ice sheet between 1949 and 2018. Detailed here are the challenges that were surmounted to obtain continuous ice cores suitable for establishing a chronological record and that would provide material for laboratory analyses.

Gertner also describes the evolution of technology that expanded our knowledge of the physical properties of Greenland's ice and enhanced our understanding of its dynamics. He communicates the importance of these accomplishments through candid observations from those involved.

Without judgment or comment, Gertner provides details worthy of philosophical reflection about the influence that military pursuits have had on the ability to conduct research in remote settings and how we value science. In the book's closing chapters, he discusses global warming, describing the controversy of whether rapid climate change has ever happened or even could happen. He also articulates its staggering implications.

Greenland and its ice will remain a place rich with opportunities for research and investigation, and Gertner's excellent book is a must-read for those who are curious about the history of exploration and the pursuit of science there.

The Ice at the End of the World: An Epic Journey into Greenland's Buried Past and Our Perilous Future, Jon Gertner, Random House, 2019, 445 pp.

For a full-length review of *The Ice at the End of the World*, see "A remote region, revealed," *Science* **364**, 1241 (2019).

10.1126/science.aba0957

[1]The reviewer is an associate editor at *Science Immunology*. Email: abalasubramani@aaas.org [2]The reviewer is the Letters editor at *Science*. Email: jsills@aaas.org [3]The reviewer is a senior editor at *Science*. Email: bwible@aaas.org [4]The reviewer is a senior editor at *Science*. Email: cash@science-int.co.uk [5]The reviewer is a publications assistant at *Science*. Email: hstewart@science-int.co.uk [6]The reviewer is a senior editor at *Science*. Email: phines@aaas.org [7]The reviewer is an eighth-grade student at Takoma Park Middle School, Silver Spring, MD 20912, USA. [8]The reviewer is a senior editor at *Science Translational Medicine*. Email: ynusinov@aaas.org [9]The reviewer is a senior editor at *Science*. Email: lzahn@aaas.org [10]The reviewer is an associate editor at *Science*. Email: mfunk@aaas.org [11]The reviewer is an associate editor at *Science*. Email: sscanlon@science-int.co.uk [12]The reviewer is a news reporter at *Science*. Email: acho@aaas.org [13]The reviewer is the executive editor of *Science*. Email: mbradfor@aaas.org [14]The reviewer is a senior editor at *Science*. Email: mlavine@aaas.org [15]The reviewer is a senior editorial fellow at *Science*. Email: asugden@science-int.co.uk [16]The reviewer is at Accenture Federal Services, Arlington, VA 22203, USA. Email: nathaniel.travis@gmail.com [17]The reviewer is at the Department of Earth and Planetary Sciences, University of California, Davis, Davis, CA 95616, USA, and the Department of Geoscience, Aarhus University, 8000 Aarhus, Denmark. Email: wglassley@ucdavis.edu

PLANETARY SCIENCE

Close-up view of an active asteroid

Particle ejections from Bennu could shape the object and send material into space

NASA's OSIRIS-REx spacecraft is going to collect a sample from the surface of asteroid Bennu in 2020 and return it to Earth in 2023.

By **Jessica Agarwal**

Most asteroids are irregularly shaped celestial bodies tens of kilometers in size or smaller that orbit the Sun inside the orbit of Jupiter. They are mostly rocky and have a wide range of compositions. This distinguishes them from comets, which originate further out in the solar system and contain a large fraction of ice that turns into gas when heated by the Sun and leads to the formation of the characteristic dust tail. On page 1217 of this issue, Lauretta *et al.* (*1*) found centimeter-sized debris ejected from the surface of the likely ice-free asteroid 101955 Bennu in images taken by NASA's OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft. The cause of this particle ejection remains somewhat puzzling, but the observations show that asteroids are far from being inert bodies (see the figure).

The notion that asteroids may not be entirely inert came with the 1996 discovery of a comet-like dust tail following the asteroid 7968 Elst-Pizarro (*2*, *3*). Roughly two dozen objects in asteroidal orbits since then were found to be temporarily ejecting dust in large enough quantities to be observable with Earth-based telescopes. Multiple causes have been identified, but the accepted main sources are sublimation of subsurface water ice (*4*) and collisions with other asteroids (*5*). Break-up or internal reconfiguration due to fast rotation may also lead to dust emission. Processes related to electrostatic charging or thermal effects are possible but less likely to induce detectable activity from Earth, unless an object is very close to the Sun (*6*).

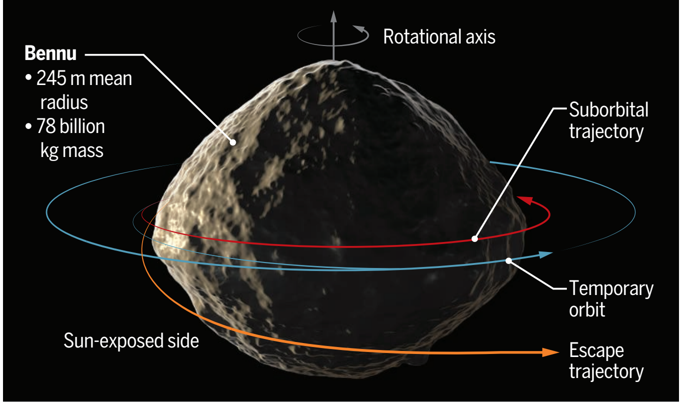
Asteroid 101955 Bennu has a diameter of ~500 m and the shape of a "spinning top"—roughly spherical, but somewhat pointed near the poles and bulgy in the equatorial zone (*7*). Bennu is categorized as a near-Earth asteroid (NEA) because it is located between 0.9 and 1.4 times the distance between Earth and the Sun. Bennu is likely composed of materials similar to a certain type of carbon-rich meteorites found on Earth called the CM chondrites (*8*). Like most asteroids of its size, Bennu is likely a "rubble pile," a collection of boulders held together mainly by their own gravity and with half of the volume empty (*7*).

The main purpose of NASA's OSIRIS-REx mission is to obtain a sample from the surface of Bennu (in 2020) and return it to Earth for analysis (in 2023). Since the beginning of 2019, the spacecraft has been thoroughly investigating the asteroid in order to pick and characterize the sampling site (*9*). Lauretta *et al.* found in images from the navigation camera particles traveling in the near environment of Bennu recorded as bright points against the dark sky. The authors identified three major ejection events in early 2019. About 100 centimeter-scale particles per event left the surface from specific locations and at specific times. The three source locations do not seem to have much in common or to be geologically different from the rest of the surface. But all three events took place in the local afternoon of Bennu. The authors also identified a background population of particles. These particles had a variety of trajectories from many different source regions and ejection times. Some particles were launched into interplanetary space, whereas others eventually fell back to Bennu's surface.

Lauretta *et al.* rule out two "standard" processes for asteroidal particle ejection: ice sublimation and rotational disintegration. Electrostatic lofting of particles from the surfaces of airless bodies was inconsistent with their observations. The authors identify micrometeorite impacts, the loss of water molecules embedded in silicate minerals, thermal stress fracturing of boulders, or a combination of these as possible causes of particle ejection. On Benn, boulder surfaces have 100 K temperature changes at a centimeter-depth scale and on

Unexpected particle ejection

Several particle ejection events from the surface of the asteroid Bennu were observed. The events all occurred during the local afternoon, possibly because of thermal cracking and higher rates of micrometeorite impact. The particles can be ejected into space, orbit the asteroid for a short time, or fall back onto the surface. NASA's OSIRIS-REx spacecraft is going to collect a sample from the surface of asteroid Bennu in 2020 and return it to Earth in 2023.



Max Planck Institute for Solar System Research, Göttingen, Germany.
Email: agarwal@mps.mpg.de

GRAPHIC: C. BICKEL/*SCIENCE*; ILLUSTRATION: NASA'S GODDARD SPACE FLIGHT CENTER

time scales of the 4.3-hour rotation period. Both the meteorite flux and the temperature at 2 cm depth should peak in the local afternoon, which lines up with the timing of the major particle ejections.

The discovery of particle ejections is surprising and may suggest that all asteroids are active at some level. However, previous spacecraft that have visited asteroids have not documented this type of event. Key answers may come from the Hayabusa2 mission by the Japan Aerospace Exploration Agency to the asteroid Ryugu (*10*), which resembles Bennu in terms of shape, density and orbit but likely is composed of different material. Like OSIRIS-REx, Hayabusa2 is characterizing the asteroid in great detail in order to bring back a sample of it in 2020. The presence or absence of activity will provide some additional constraints on the low-level activity of small asteroids.

The Bennu observations leave open a question of how to predict such low-level activity of the many millions of asteroids not targeted for missions. Estimating the amount of debris that is released into interplanetary space is clearly important but very challenging. Depending on the driving process for Bennu's ejections, there may or may not be a continuum between it and "classically" active asteroids. The activity on Bennu shows that an apparently inactive asteroid can harbor a complex dynamic of debris reimpacting the surface or feeding the interplanetary dust cloud. This process may have implications for Bennu's evolution. A similar observation was made at comet 67P/Churyumov-Gerasimenko by the European Space Agency's Rosetta mission, which found small outbursts of gas and debris emission through the comet's entire perihelion passage (*11*). Earth-based telescopes did not pick up this activity (*12*). Isolating the detailed processes across different scales that drive activity on asteroids and comets is important to understand the evolution of these bodies and how they supply dust to interplanetary space. ■

REFERENCES AND NOTES

1. D. S. Lauretta *et al*., *Science* **366**, eaay3544 (2019).
2. E. W. Elst *et al*., *IAU Circular* 6456 (1996).
3. H. H. Hsieh, D. C. Jewitt, Y. R. Fernández, *Astron. J.* **127**, 2997 (2004).
4. H. H. Hsieh, D. Jewitt, *Science* **312**, 561 (2006).
5. M. Ishiguro *et al*., *Astrophys. J.* **741**, L24 (2011).
6. D. Jewitt *et al*., in *Asteroids IV*, P. Michel *et al*., eds. (Univ. Arizona Press, 2015).
7. D. S. Lauretta *et al*., *Nature* **568**, 55 (2019).
8. D. N. DellaGiustina *et al*., *New Astron.* **3**, 341351 (2019).
9. D. S. Lauretta *et al*., *Space Sci. Rev.* **212**, 925 (2017).
10. S. Watanabe *et al*., *Science* **364**, 268 (2019).
11. J.-B. Vincent *et al*., *Mon. Not. R. Astron. Soc.* **462** (suppl.1), S184 (2016).
12. H. Böhnhardt *et al*., *Mon. Not. R. Astron. Soc.* **462**, S376 (2016).

ACKNOWLEDGMENTS

This work was supported by the European Research Council Starting Grant 757390 "CAstRA."

10.1126/science.aaz7129

STEM CELLS

Lymphatic vessels as a stem cell niche

Reciprocal signaling between lymphatic vessels and hair follicle stem cells drives skin regeneration

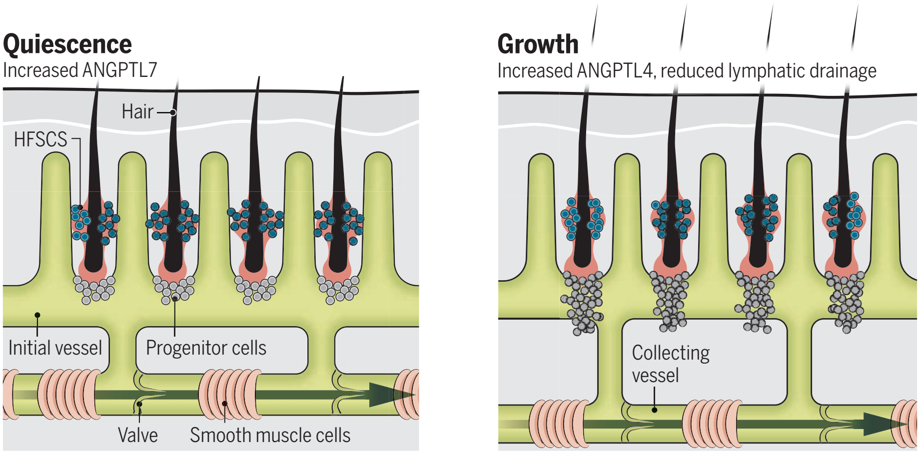
By **Natasha L. Harvey**

Lymphatic vessels play crucial roles in tissue fluid homeostasis, immune cell trafficking, and dietary lipid absorption. An extensive network of lymphatic vessels beneath the epidermis provide a highway for the transport of immune cells that patrol the skin to lymph nodes, where antigens are presented to the immune system for surveillance (*1*). On page 1218 of this issue, Gur-Cohen *et al.* (*2*) reveal that in addition to regulating interstitial fluid balance and directing immune cell traffic, lymphatic vessels in mouse skin comprise an important stem cell niche. This intriguing work demonstrates that superficial lymphatic vessels located in the dermis share an intimate and dynamic spatial association with hair follicle stem cells (HFSCs) and that signaling between these two compartments coordinates hair follicle cycling and skin regeneration.

Stem cell niches are specialized microenvironments within tissue, often comprising multiple cell types, that support stem cell activity. Extensive evidence demonstrates that blood vessels are an integral component of stem cell niches in tissues, including in bone marrow, muscle, and brain (*3*). However, the role of lymphatic vessels in regulating stem cell activity has, until now, remained unexplored. Gur-Cohen *et al.* used recent advances in imaging technology to reveal that superficial, initial lymphatic vessels, which form a stereotypical pattern throughout skin, envelope hair follicles and share particularly close contact with HFSCs. Initial lymphatic vessels are specialized for fluid absorption. Their characteristics include prominent expression of lymphatic vessel endothelial hyaluronan acid receptor 1 (LYVE1) on lymphatic endothelial cells, which line the vessel (*4*), and relatively loose, button-like endothelial cell junctions that enable the entry of fluid and cellular traffic to this part of the lymphatic vasculature (*5*). Fluid, macromolecules, and immune cells taken up by the initial lymphatics are drained into precollecting and collecting lymphatic vessels, which exhibit a distinct identity and are specialized to return lymph to the bloodstream. Collecting vessels have tighter, zipper-like endothelial cell junctions, contain valves, and are surrounded by vascular smooth

The hair follicle stem cell niche

During quiescence, hair follicle stem cells (HFSCs) produce factors, including angiopoietin-related protein 7 (ANGPTL7), that maintain contact between HFSCs and initial lymphatic vessels and promote lymphatic vessel drainage. The shift of HFSCs to the production of ANGPTL4 results in dissociation of lymphatic vessels from the HFSC niche, reduced lymphatic drainage, and hair growth.



GRAPHIC: V. ALTOUNIAN/SCIENCE

muscle cells, which together facilitate the transport function of these vessels (*4*, *5*).

The close proximity of HFSCs to initial lymphatic vessels raised the possibility that signaling between stem cells and lymphatic endothelial cells might regulate stem cell activity, lymphatic vessel function, or both. Gur-Cohen *et al.* found that disrupting lymphatic vessel integrity in mouse models promoted hair follicle proliferation and hair regeneration. These data suggested that signals from lymphatic endothelial cells normally render HFSCs quiescent (noncycling).

To explore whether the association between initial lymphatics and HFSCs was temporally regulated, lymphatic vessel pattern and function were assessed during the cycle of hair growth and quiescence. These analyses demonstrated that lymphatic vessel proximity to the bulge region of the follicle, where HFSCs reside, changed during cyclical phases of hair growth and quiescence: Initial lymphatic vessels were closely associated with HFSCs during the resting state and dissociated during the proliferation and growth phase. Lymphatic vessels also had reduced drainage capacity during the proliferative phase of the hair cycle. It will be intriguing in future studies to understand whether the reduced drainage occurs as a result of a spatial change in tissue architecture or a functional change in lymphatic absorptive capacity.

To investigate the identity of factors likely to be important for communication between the HFSC niche and lymphatic vessels, Gur-Cohen *et al.* profiled the gene expression of quiescent and proliferative HFSCs. Comparison of these datasets revealed a number of factors, some of which had previously been implicated in vascular remodeling, which were prominent either during follicle quiescence—in particular, angiopoietin-related protein 7 (ANGPTL7)—or during HFSC proliferation—particularly, ANGPTL4. Prolongation of ANGPTL7 expression through the growth phase of the hair cycle maintained the close association of lymphatic vessels with the HFSC compartment, resulting in maintenance of stem cell quiescence. Conversely, induced expression of ANGPTL4 during quiescence resulted in premature dissociation of lymphatic vessels from the niche and induction of stem cell proliferation. The expression of the *Angptl4* gene in mouse endothelial cells is important for maintaining separation of the blood and lymphatic vascular compartments in the postnatal mouse intestine (*6*). This suggests that ANGPTL4 provides a repulsion signal to lymphatic endothelial cells that is important in multiple tissue contexts.

Gur-Cohen *et al.* found that reducing ANGPTL7 expression induced hyperplastic follicles that entered the growth phase in an asynchronous manner across the skin. Moreover, lymphatic vessels were dilated and exhibited reduced drainage capacity in this scenario. Analysis of factors expressed by lymphatic endothelial cells in different phases of the hair cycle revealed that putative receptors for ANGPTL4 and ANGPTL7 are present on initial lymphatics. In addition, WNT signaling inhibitors that restrict HFSC activity and hair growth were produced by lymphatic vessels during HFSC quiescence.

Together, these data suggest that factors secreted by stem cells directly affect lymphatic vessel structure and/or function and vice versa (see the figure). It will be interesting to determine whether ANGPTL4 and ANGPTL7 have the same impact on lymphatic vessels in other tissues and to identify the receptors that transduce ANGPTL signals in lymphatic endothelial cells. It will also be fascinating to understand the mechanisms by which these factors regulate lymphatic vessel absorptive capacity, particularly in the settings of inflammation and disease. This may result from active remodeling of cell junctions. The possibility also exists that mechanical signals, as a result of changes in interstitial volume due to lymphatic vessel remodeling, affect stem cell activity; this will also be important to explore.

Additional recent studies have described the intimate association of dermal lymphatic vessels with the HFSC niche (*7*, *8*). Different mouse models were used in these studies to investigate the impact of increased or decreased lymphatic vessel density on HFSC activity, but all concluded that signaling between these two cellular compartments is important for the coordination of hair growth in the skin. Further studies exploring the link between lymphatic vessels and stem cell activity during development and in disease could identify new axes that can be targeted to enhance tissue repair and regeneration. ■

REFERENCES AND NOTES

1. T. Tammela, K. Alitalo, *Cell* **140**, 460 (2010).
2. S. Gur-Cohen *et al.*, *Science* **366**, 121 (2019).
3. S. Rafii, J. M. Butler, B.-S. Ding, *Nature* **529**, 316 (2016).
4. T. Mäkinen *et al.*, *Genes Dev.* **19**, 397 (2005).
5. P. Baluk *et al.*, *J. Exp. Med.* **204**, 2349 (2007).
6. F. Bäckhed, P. A. Crawford, D. O'Donnell, J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 606 (2007).
7. D. Peña-Jimenez *et al.*, *EMBO J.* **38**, e101688 (2019).
8. S. Y. Yoon *et al.*, *PLOS ONE* **14**, e0220341 (2019).

Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, Australia.
Email: natasha.harvey@unisa.edu.au

10.1126/science.aaz8780

MOLECULAR BIOLOGY

Folding unpredicted

Unexpected topology is the key to glutamate receptor gating in neurotransmission

By **Jochen Schwenk** *and* **Bernd Fakler**

Cornichon homologs (CNIHs) are a family of highly conserved small membrane proteins (140 to 160 amino acids in length) that serve as shuttle(s) for the export of proteins from the endoplasmic reticulum (ER) and/or as auxiliary subunits that control the building and gating of AMPA-type glutamate receptors (AMPARs) in the brain. According to predictions from databases and algorithms, CNIHs exhibit a three-transmembrane (TM) domain topology with the protein's amino terminus facing the cytoplasm. On page 1259 of this issue, Nakagawa pre-sents the cryo–electron microscopy (cryo-EM) structure of a CNIH protein in complex with a glutamate receptor (*1*). The results prove database predictions wrong and provide a structural key to the diverse functions of the cornichon family of proteins.

"...predictions from databases and algorithms are good, but rigorous experimental data are better."

The founding member of this family, the classical cornichon (CNIH1 in mammals), was described ~20 years ago in the fruit fly *Drosophila melanogaster* (*2*). It serves as a cargo receptor required for efficient ER export of the growth factor gurken (*3*). Mechanistically, cornichon recognizes and binds its cargo and transfers it to the secretory pathway via recruitment into coat protein complex II (COPII)–coated vesicles (*3*). This action was later detailed in yeast, where the cornichon homolog Erv14 (ER-derived vesicles protein 14) is able to accept a broad

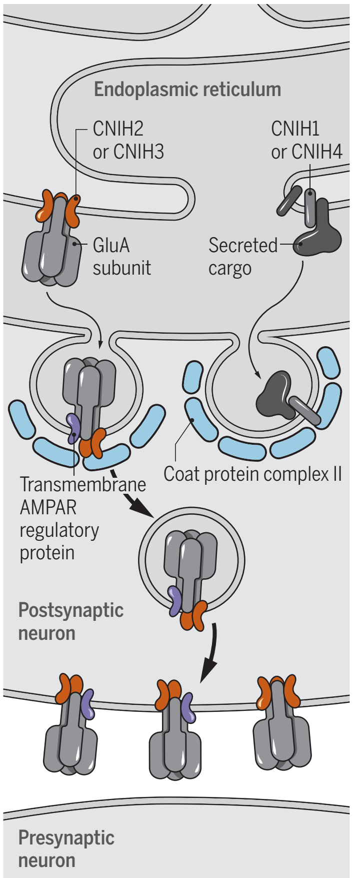
Institute of Physiology II, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
Email: bernd.fakler@physiologie.uni-freiburg.de

Functions of cornichon proteins

Cornichon homolog 2 (CNIH2) and CNIH3 control endoplasmic reticulum (ER) export and gating of AMPA-type glutamate receptors (AMPARs, comprising GluA subunits). CNIH1 and CNIH4 are nonspecific cargo receptors.

At the ER
CNIH2 and CNIH3 specifically bind to AMPARs through interactions with the amino terminus and the peptide extensions of the cytoplasm-facing region of these CNIHs.

At the synapse
CNIH2 and CNIH3 are auxiliary subunits of AMPARs that stabilize the channels in the open state and thereby enhance signal transduction.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

spectrum of cargo proteins and cargo recognition is profoundly affected by the length of the cargo's TM domains rather than by a specific amino acid sequence motif (or motifs) (*4*).

In this context, it came as a surprise that in mammals, two of the four cornichon homologs, CNIH2 and CNIH3, are specialized toward AMPARs, the most abundant excitatory neurotransmitter receptor in the brain, by recognizing the pore-forming GluA1–4 subunits as exclusive "cargo" or interaction partners (*5*). However, the CNIH2-GluA and CNIH3-GluA interactions differ from the CNIH1-cargo interaction in two characteristic ways. One is that CNIH2 and CNIH3 can only bind to their cargo in the ER after GluAs have been assembled into tetramers, which, in neurons, requires an ER-based assembly machinery (*6*). Another is that CNIH2 and CNIH3 do not release their cargo, but rather remain tightly associated and travel with the AMPARs to the plasma membrane, where they have been found in both synapses and extrasynaptic sites (*7*–*9*) (see the figure).

Thus, in addition to their role as cargo exporters, CNIH2 and CNIH3 are auxiliary subunits of AMPARs that have a profound impact on the gating properties of the receptor channels: Both CNIH2 and CNIH3 stabilize the AMPAR channels in the open or conducting state(s) by slowing the deactivation (channel closing after agonist removal) and desensitization (channel closing with agonist bound) processes (*5*). In excitatory synapses, such prolonged opening enhances the influx of Na^+ (and Ca^{2+}) into the postsynapse, which, through increased and longer-lasting depolarization of the postsynaptic neuron, leads to more reliable propagation of the electrical signal. This was directly observed in hippocampal neurons (*7*).

For all distinct cornichon functions, the characterization of CNIH3-GluA2 complexes by Nakagawa provides interesting insights. The most intriguing and unpredicted features of this cryo-EM structure are the topology and overall folding of CNIH3: The protein displays four TM domains with amino and carboxyl termini facing the ER lumen (which becomes the extracellular side of the plasma membrane). Most parts of the protein are buried in the membrane; only the linker between TM domains 3 and 4, and parts of TM domains 1 and 2, extend into the cytoplasm. This unexpected topology may be initiated by a 12–amino acid stretch at the amino terminus of CNIH1–3 that is highly conserved among mammals, flies, and worms and operates as a signal peptide that is not cleaved off [therefore called "uncleavable membrane inserting peptide" (UMIP) by the author]. The amino-terminal half of the UMIP contains three conserved phenylalanine residues that make close contact with the TM domains of the GluA subunits.

Projection of the primary sequence(s) onto the CNIH3 structure highlights a 4.5-turn α-helical extension at the cytoplasmic side of TM2 as the major difference between the GluA-specific CNIH2 and CNIH3 and the cargo receptors CNIH1 and CNIH4. This extension, together with the helical domain of TM1 and the (structurally unresolved) linker, represents the hallmark feature that distinguishes the CNIHs from the other inner core constituents of AMPARs, the TARPs (transmembrane AMPAR regulatory proteins), and GSG1L (germ cell–specific gene 1–like protein) (*10*, *11*). TARPs and GSG1L also modulate channel gating and comprise four TM segments, but their "extensions" are directed toward the extracellular side of the membrane where they can contact the ligand-binding domains (LBDs) of the GluA subunits during the conformational changes that drive channel gating after agonist binding (*11*). Because CNIH2 and CNIH3 lack contacts with the LBDs, Nakagawa suggests that structural rearrangements induced by CNIH3 in the "filter" region of the receptor channel likely explain the pronounced effects on channel gating. The filter contains a particular site that controls Ca^{2+} permeability and may influence the open-state configuration of the receptor channel (*12*). Whether these rearrangements are relevant for AMPAR gating remains unclear, but they are strongly reminiscent of the "selectivity-filter gate" that opens (and closes) various types of K^+ channels in response to distinct ligands and pharmacological agents (*13*).

The CNIH3-GluA2 complexes are rarely (if at all) observed in the rodent brain. Most CNIH-containing AMPARs (a total of ~80% of all AMPARs) have both CNIH and TARP proteins co-assembled in a 2:2 stoichiometry, as estimated by quantitative proteomics (*14*) and confirmed by the cryo-EM analysis of native AMPARs (*15*). Most of the successfully resolved structures of native AMPARs (*15*) have contained two TARP subunits and two subunits with four densities in the membrane plane; the CNIH3-GluA2 structure from Nakagawa suggests that these may be CNIH2 and/or CNIH3 proteins.

The unexpected structure of CNIH3 and its distinction from the TARPs will provide a blueprint for coming analyses of its role in gating and trafficking of AMPARs, but also in studying the ER export processes mediated by the CNIH family. Finally, it should not be forgotten that predictions from databases and algorithms are good, but rigorous experimental data are better. ■

REFERENCES AND NOTES

1. T. Nakagawa, *Science* **366**, 1259 (2019).
2. S. Roth *et al.*, *Cell* **81**, 967 (1995).
3. C. Bökel *et al.*, *Development* **133**, 459 (2006).
4. Y. Herzig *et al.*, *PLOS Biol.* **10**, e1001329 (2012).
5. J. Schwenk *et al.*, *Science* **323**, 1313 (2009).
6. J. Schwenk *et al.*, *Neuron* **104**, 680 (2019).
7. S. Boudkkazi *et al.*, *Neuron* **82**, 848 (2014).
8. B. E. Herring *et al.*, *Neuron* **77**, 1083 (2013).
9. A. S. Kato *et al.*, *Neuron* **68**, 1082 (2010).
10. B. Herguedas *et al.*, *Science* **364**, 353 (2019).
11. E. C. Twomey *et al.*, *Nature* **549**, 60 (2017).
12. I. D. Coombs *et al.*, *Nat. Commun.* **10**, 4312 (2019).
13. M. Schewe *et al.*, *Science* **363**, 875 (2019).
14. J. Schwenk *et al.*, *Neuron* **74**, 621 (2012).
15. Y. Zhao *et al.*, *Science* **364**, 355 (2019).

ACKNOWLEDGMENTS

We thank members of the Fakler lab for discussions.

10.1126/science.aaz8642

Oil and gas exploration fragments boreal forest in Canada.

ECOLOGY

Lasting signature of forest fragmentation

Animal communities that endured historical environmental upheavals are less sensitive to modern ones

By **Anna Hargreaves**

A universal truth of ecology is that field experiments never unanimously support theory. This is not (always) because ecological theory is poorly developed or experiments poorly executed, but because ecology is a complex science dealing with variation at every biological level from individuals to biomes. When exceptions are the rule, explaining variation in responses among taxa and locations becomes the goal, particularly for theory that informs conservation. On page 1236 of this issue, Betts *et al.* (*1*) contribute to a particularly important debate: why the biological effects of forest fragmentation are so variable among species and places. They present evidence that historical deforestation (from glaciation, fires, hurricanes, or anthropogenic clearing) yielded communities that are more robust to modern forest fragmentation (from logging, burning, or development).

Department of Biology, McGill University, Montreal, Canada. Email: anna.hargreaves@mcgill.ca

Humans are rapidly converting natural habitats (*2*), and the effects come in two flavors: loss and fragmentation. The negative impact of habitat loss is undeniable: Populations decline; species go locally extinct. Indeed, after direct exploitation, habitat loss is the primary cause of past extinctions (*3*) and a serious threat to modern biodiversity (*4*). Fragmentation refers to the spatial arrangement of the remaining habitat after some is lost, in particular the frequency of habitat edges. The relative impact of fragmentation after accounting for habitat loss has been the subject of raging debate since the early 2000s (*5*–*8*). Whereas landscape-level tests are still too rare to reach robust conclusions, case studies show that fragmentation per se can be devastating for some taxa in some places but surprisingly positive for others (*9*). This variation has not yet been convincingly explained.

Although the ecological effects of fragmentation have been studied extensively (if not conclusively), the evolutionary implications have not. Short-term evolution after fragmentation has been documented, albeit rarely, mostly involving changes in dispersal or mating systems (*10*). Such evolution could help populations survive in fragmented landscapes (for example, evolution of increased dispersal ability could reduce perceived fragmentation) but could also create evolutionary traps (for example, evolution of decreased dispersal might doom populations to stay in disappearing patches) (*11*). The jury is out on whether evolution will be rapid enough and in the right direction to help species cope with modern fragmentation. Even less is known about the long-term evolutionary effects of historical fragmentation.

Betts *et al.* tackle these gaps through the lens of extinction filters. The extinction filter hypothesis proposes that historical exposure to a stressor filters a community to species that can cope with that particular stressor (*12*). This hypothesis has most famously been applied to ancient human hunting, proposing that areas with a longer history of human exploitation have faunas that are less vulnerable to modern exploitation (*13*). Extinction filters can operate evolutionarily if species adapt to tolerate the disturbance, or ecologically if sensitive species are lost during disturbance. Betts *et al.* predict that animal communities in forests historically prone to edge-creating disturbances should be more resilient to modern fragmentation, either because species have adapted to edge effects or because sensitive species have already been purged. Consistent with the extinction filter hypothesis, Betts *et al.* find fewer species that specialize on interior forest habitat and less edge avoidance in forests with a history of severe disturbance in the past 10,000 years.

These results partly rest on correctly identifying areas with historical edge-creating disturbances. The authors define such areas as those subject to regular hurricanes, forest fires, or glaciation (all binary variables), or recent human deforestation measured as the amount of intact forest predicted to exist by climate models that was missing in the year 2000. The time scale of these disturbances is clearly highly variable, and their primary effect is undoubtedly forest loss. But because even glaciers have uneven edges and storms leave pockets untouched, it is reasonable to assume that these events created increased edge effects and some degree of fragmentation as well.

Taken together, these disturbances leave a band of low-disturbance forest concentrated around the equator. Accordingly, Betts *et al.* found that tropical forest communities are more sensitive to edge effects, and that the proportion of forest-core specialists in-

creases from high to low latitudes, contrary to results of a recent review that found no difference in sensitivity between tropical and temperate taxa (*9*).

Support for the extinction filter hypothesis can be interpreted optimistically or pessimistically for conservation, depending on whether filtering is evolutionary or ecological. If historically disturbed communities can successfully adapt to disturbance, we might hope that future evolution will rescue at least some of the habitat specialists currently threatened by fragmentation. Betts *et al.* found that past human deforestation—arguably the disturbance with the shortest evolutionary time scale—was a much weaker predictor of edge sensitivity than natural disturbances. Perhaps evolution simply has not had time to mitigate the ill effects of human activity.

However, the measure of historical human deforestation used by Betts *et al.* only includes areas that were still deforested as of 2000. The true extent of human forest use in the past 10,000 years is controversial, particularly in the Americas, where European contact in 1492 wiped out 90% of indigenous inhabitants (numbering in the millions) in as little as a generation (*14*). Recent estimates suggest this "great dying" led to the reverting of cleared areas to forest (*15*). If species could evolve to cope with human forest fragmentation on evolutionary time scales (extinction filtering via adaptation), we might expect modern deforestation to overpredict sensitivity in areas where forests regrew before 2000. In fact, Latin American forests make up most of the low-disturbance communities in the analysis and seem disproportionately sensitive to fragmentation. If extinction filtering results instead from purging of sensitive species, Betts *et al.*'s results suggest a grim future for tropical forest specialists that are rapidly running out of intact forest in which to seek refuge (*2*). ■

REFERENCES AND NOTES

1. M. G. Betts *et al.*, *Science* **366**, 1236 (2019).
2. F. Taubert *et al.*, *Nature* **554**, 519 (2018).
3. J. K. Szabo, N. Khwaja, S. T. Garnett, S. H. M. Butchart, *PLOS ONE* **7**, e47080 (2012).
4. S. L. Maxwell, R. A. Fuller, T. M. Brooks, J. E. M. Watson, *Nature* **536**, 143 (2016).
5. L. Fahrig, *Annu. Rev. Ecol. Evol. Syst.* **34**, 487 (2003).
6. R. J. Fletcher Jr. *et al.*, *Biol. Conserv.* **226**, 9 (2018).
7. E. I. Damschen *et al.*, *Science* **365**, 1478 (2019).
8. L. Fahrig *et al.*, *Biol. Conserv.* **230**, 179 (2019).
9. L. Fahrig, *Annu. Rev. Ecol. Evol. Syst.* **48**, 1 (2017).
10. P. O. Cheptou, A. L. Hargreaves, D. Bonte, H. Jacquemyn, *Philos. Trans. R. Soc. B* **372**, 20160037 (2017).
11. K. Parvinen, *Evol. Ecol. Res.* **9**, 619 (2007).
12. A. Balmford, *Trends Ecol. Evol.* **11**, 193 (1996).
13. S. T. Turvey, S. A. Fritz, *Philos. Trans. R. Soc. B* **366**, 2564 (2011).
14. N. Nunn, N. Qian, *J. Econ. Perspect.* **24**, 163 (2010).
15. A. Koch, C. Brierley, M. M. Maslin, S. L. Lewis, *Quat. Sci. Rev.* **207**, 13 (2019).

10.1126/science.aba1103

IMMUNOTHERAPY

Ushering along B cells to neutralize HIV

Progress in staged immunizations designed to elicit a vaccine response is reported

By **Amanda Agazio** *and* **Raul M. Torres**

Although vaccines are a great achievement in medicine, HIV, with its extraordinary dynamic diversity, is not restrained by classic vaccine approaches. In 2009, after the RV144 vaccine trial results revealed suboptimal HIV protection, effort refocused on developing vaccines able to elicit antibodies that can protect against a breadth of HIV genetic variants. Ten years on, such HIV broadly neutralizing antibodies (bnAbs) are known to display unusual features compared with typical antibodies. These features almost certainly impede eliciting bnAb generation with a vaccine. On pages 1215 and 1216 of this issue, Saunders *et al.* (*1*) and Steichen *et al.* (*2*), respectively, use custom-designed HIV envelope (Env) proteins as immunogens in animal models to promote unusual antibody features needed for neutralization and to recruit rare bnAb-precursor B cells into antibody responses. These studies demonstrate progress in eliciting antibodies with the potential to provide a breadth of HIV neutralizing activity.

In the past decade, numerous technological advances have greatly facilitated the identification and characterization of bnAbs from HIV-infected individuals. These include single cell–based approaches to identify HIV-specific B cells, molecularly clone their antibody receptors, and rapidly screen for specificity and neutralizing activity. When coupled with bnAb-Env structural analyses, the development of stable Env proteins and computational design have led to custom-designed Env proteins that are able to recruit rare B cells into an antibody response and have fueled progress toward an HIV vaccine (*3*, *4*). Many hundreds of bnAbs have now been isolated from HIV-infected individuals and characterized to reveal that bnAbs recognize one of a few conserved regions, or epitopes, on Env and neutralize by preventing HIV association with the CD4 T cell receptor (the target of infection) and blocking virion-cell fusion.

Critically, bnAb recognition of these neutralizing epitopes requires one or more unusual features not normally observed in antibodies elicited by other pathogens. For example, the immunoglobulin (Ig) heavy chain region that contacts Env [heavy chain complimentary-determining region 3 (HCDR3)] is often much longer in bnAbs compared with other antibodies. In the course of a normal antibody response, Ig genes are mutated to increase antibody affinity for the targeted pathogen (affinity maturation), and bnAbs are also unusual in that they display a high frequency of mutations, including rare mutations that are difficult to generate but important for broad recognition of HIV variants. Longitudinal studies of HIV-infected individuals have characterized the coevolution of HIV Env and antibody responses to reconstruct the trajectory, or lineage, of a bnAb from the original unmutated antibody, highlighting the mutations that are important for neutralization (*5*).

Together, these studies motivated an Env structure–based approach in which distinct engineered immunogens, through sequential immunizations, drive the maturation of Env-recognizing B cells to produce bnAbs. An initial immunogen would elicit (and numerically expand) naïve B cells expressing unmutated antibodies into an antibody response with the potential to further develop into a bnAb. Immunization with a second immunogen would select responding B cells with specific antibody mutations that are important for recognition of the neutralizing epitope (see the figure). In this manner, serial immunizations would usher B cells with specific somatic mutations to ultimately generate a B cell population that secretes antibodies capable of providing a wide breadth of neutralizing activity against HIV (*1*, *6*, *7*). Previous studies using mice with B cells engineered to express bnAbs (*4*, *8*) demonstrated feasibility and provided an impetus for pursuing this approach.

A major issue addressed by Saunders *et al.* is that bnAbs often require mutations for neutralizing activity that are not typically introduced into antibodies during the

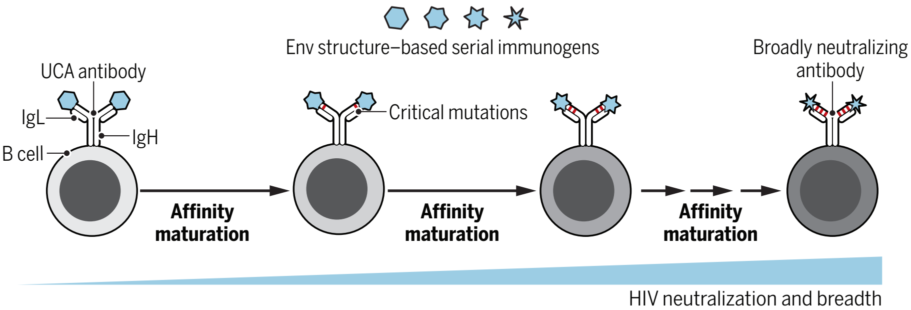
Department of Immunology & Microbiology, University of Colorado School of Medicine, 12800 East 19th Avenue, Aurora, CO 80045, USA. Email: raul.torres@cuanschutz.edu

antibody response and thus represent bottlenecks in generating bnAbs (*9*). Two previously described bnAbs, DH270 and CH235 (*10*, *11*), the lineages of which were previously characterized and shown to harbor improbable mutations that are important for HIV-1 neutralization (*9*), were studied by Saunders *et al.* On the basis of Env-DH270 lineage bnAb structures, an Env immunogen was designed that bound the inferred unmutated common ancestor (UCA) antibody with submicromolar affinity and, critically, also bound an affinity-maturation intermediate bnAb harboring rare mutations with greater than fourfold increased affinity. After six immunizations with a nanoparticle-Env immunogen, physiologically relevant neutralizing activity of modest breadth was achieved in mice harboring B cells engineered to express the DH270-UCA bnAb precursor antibody. This Env immunogen recruited DH270-UCA–expressing B cells eliciting HIV-1 neutralizing antibodies from a natural B cell repertoire.

Steichen *et al.* similarly achieve an important goal in the development of an effective HIV-1 vaccine investigating the BG18 bnAb lineage that leads to a potent bnAb recognizing the V3 glycan neutralizing epitope on Env (*12*). However, this study focuses on the available B cell antibody specificities in a healthy individual. Because each of the hundreds of millions of B cells in the body express a unique antibody specificity, the frequency of precursor B cells that can recognize a bnAb-eliciting immunogen is important to determine when considering which precursor bnAb lineage to target (*13*). Although the BG18 bnAb sequence is 30% mutated from the inferred UCA, this bnAb lacks nucleotide deletions or insertions that are considered more difficult to elicit through affinity maturation. Informed by the BG18-Env

Staged HIV-1 immunization

Serial immunizations with distinct immunogens of the Env epitope of HIV-1 favor affinity maturation of antibodies toward broadly neutralizing activity in mice and macaques.



Env, envelope; IgH, immunoglobulin heavy chain; IgL, immunoglobulin light chain; UCA, unmutated common ancestor.

that incorporated rare mutations critical for neutralization, an important advance on the path to bnAb generation.

In additional experiments, inferred CH235-UCA antibody genes were again used to produce mice that, when immunized with Env immunogen, generated HIV-1 neutralizing antibodies also carrying CH235 rare mutations. Inducing physiologically relevant HIV-1 neutralization in mouse models is noteworthy, although tempered, in these studies because the mice were engineered to harbor supraphysiological (10 to 15%) initial precursor B cell populations that were capable of responding. However, rhesus macaques immunized with the CH235-eliciting Env immunogen elicited HIV-1 neutralizing antibodies after at least two immunizations. Macaques presumably harbor a physiologically relevant low frequency of immunogen-binding B cells, although this was not evaluated. Thus, the findings in nonhuman primates highlight an important hallmark in structure, Steichen *et al.* designed a precursor antibody that carried 11% of the mutations while retaining two-thirds the BG18 neutralizing breadth. On the basis of this minimally mutated antibody sequence and antibody-Env structure, next-generation DNA sequencing data from 14 healthy individuals were examined to identify Ig heavy chain–encoding sequences from each donor with similar BG18 HCDR3 length and Ig gene usage; 28 common BG18-like precursor antibodies were identified.

Common precursor antibodies were produced and used to iteratively select Env immunogens with appreciable affinities to the common BG18 precursor antibodies. These immunogens were then used to immunize mice engineered to harbor B cells expressing the BG18-UCA Ig heavy chain that could pair with different mouse Ig light chains. To better reflect physiological precursor B cell frequencies, BG18-UCA$^+$ B cells were adoptively transferred into wild-type hosts at a ratio of seven BG18-UCA$^+$ B cells per million host B cells. These mice were immunized with nanoparticle-presented Env immunogen, and rare BG18-UCA$^+$ precursor B cells were shown to respond, including their differentiation into plasma cells secreting IgG antibodies with increased affinities for the immunogen. Steichen *et al.* further used these BG18-UCA immunogens to assess reactivity among human B cells from 16 different donors and found that immunogen-reactive B cells were present at a frequency of ~0.001% within a healthy B cell antibody repertoire. These data thus document the feasibility of engineering HIV-1 immunogens that can recruit precursor bnAb B cells into an antibody response from a physiologically relevant rare subset.

Development of an HIV vaccine remains a daunting task. The studies of Saunders *et al.* and Steichen *et al.*, together with a recent report from Escolano *et al.* demonstrating the ability of a distinct HIV-1 Env immunogen to recruit precursor bnAb-expressing B cells into an appropriate antibody response in wild-type mice, rabbits, and macaques (*14*), demonstrate meaningful progress. A number of issues remain important to consider in future studies, including identifying which of the bnAb lineages to replicate in a vaccine and the ability of any new immunogen to also generate B cell memory and long-lived plasma cells, as well as better characterization of the breadth of neutralizing activity elicited. Inherent to reaching these goals will be a more careful evaluation of the manner in which immunogens are presented in a vaccine as well as the timing and mode of vaccine administration (*15*). Finally, considering that many HIV bnAbs cross-react with self-antigens, it will also be important to evaluate whether vaccine strategies targeting these lineages also promote autoimmunity. ■

REFERENCES AND NOTES

1. K. O. Saunders *et al.*, *Science* **366**, eaay7199 (2019).
2. J. M. Steichen *et al.*, *Science* **366**, eaax4380 (2019).
3. P. D. Kwong, J. R. Mascola, *Immunity* **37**, 412 (2012).
4. P. D. Kwong, J. R. Mascola, *Immunity* **48**, 855 (2018).
5. M. Bonsignori *et al.*, *Immunol. Rev.* **275**, 145 (2017).
6. B. F. Haynes *et al.*, *Cell Host Microbe* **19**, 292 (2016).
7. A. Escolano, P. Dosenovic, M. C. Nussenzweig, *J. Exp. Med.* **214**, 3 (2017).
8. J. E. Crowe Jr., *Cell* **166**, 1360 (2016).
9. K. Wiehe *et al.*, *Cell Host Microbe* **23**, 759 (2018).
10. M. Bonsignori *et al.*, *Sci. Transl. Med.* **9**, eaai7514 (2017).
11. M. Bonsignori *et al.*, *Cell* **165**, 449 (2016).
12. N. T. Freund *et al.*, *Sci. Transl. Med.* **9**, eaal2144 (2017).
13. R. K. Abbott *et al.*, *Immunity* **48**, 133 (2018).
14. A. Escolano *et al.*, *Nature* **570**, 468 (2019).
15. K. M. Cirelli *et al.*, *Cell* **177**, 1153 (2019).

ACKNOWLEDGMENTS

We acknowledge R. Pelanda for helpful comments. R.M.T is funded by the National Institutes of Health (NIH, AI136534), and A.A. was funded by the NIH (T32 AR007534).

10.1126/science.aaz8647

GRAPHIC: VERONICA FALCONIERI/*SCIENCE*

BIOCATALYSIS

Biocatalytic cascades go viral

An investigational drug targeting the HIV virus is synthesized with nine enzymes

By **Elaine O'Reilly** *and* **James Ryan**

Natural biosynthesis assembles a vast array of complex natural products starting from a limited set of building blocks, under physiological conditions, and in the presence of numerous other biomolecules. Organisms rely on the extraordinary selectivity of enzymes and their ability to operate under similar reaction conditions, meaning that these catalysts are perfectly adapted to mediate cascade reactions. In these multistep processes, the product of one biocatalytic step becomes the substrate for the next transformation (*1*–*3*). On page 1255 of this issue, Huffman *et al.* (*4*) report the development of an impressive nine-enzyme biocatalytic cascade for the synthesis of the investigational drug islatravir for the treatment of human HIV.

This study represents a partnership between scientists from Merck and Codexis. These two companies have a history of successfully collaborating to develop biocatalysts for the synthesis of important pharmaceuticals. Almost a decade ago, they developed a chemoenzymatic route for the synthesis of the type 2 diabetes drug sitagliptin (Januvia), relying on a key enzyme-catalyzed transamination with a highly engineered (*R*)-selective transaminase (*5*). The work was considered a landmark example of directed evolution and functioned to highlight the potential application of biocatalysis to revolutionize industrial chemical processes.

School of Chemistry, University College Dublin, Belfield, Dublin 4, Ireland. Email: elaine.oreilly@ucd.ie

The cascade for synthesizing islatravir was inspired by the bacterial nucleoside salvage pathway, which recycles precious nucleosides by using three key enzymes: a purine nucleoside phosphorylase (PNP), a phosphopentomutase (PPM), and a deoxyribose-5-phosphate aldolase (DERA) (see the figure). However, to achieve the synthesis of the target molecule, Huffman *et al.* required the natural nucleoside degradative cascade to run in reverse. The reversible nature of enzymes is central to the design of this cascade and is one of the important features that sets biocatalysts apart from the majority of traditional chemical catalysts.

The success of the cascade developed by the team also relied on all three enzymes accepting non-natural substrates bearing a fully substituted carbon at the C-4 position of the 2-deoxyribose ring. The authors reconstructed the reverse nucleoside salvage pathway from a PNP and PPM found in *Escherichia coli* and a DERA from *Shewanella halifaxensis*. The native *E. coli* enzymes required engineering to improve their activity. The DERA displayed existing high activity and stereoselectivity for the formation of the desired sugar phosphate enantiomer, but it required engineering to improve its ability to operate at high substrate concentration.

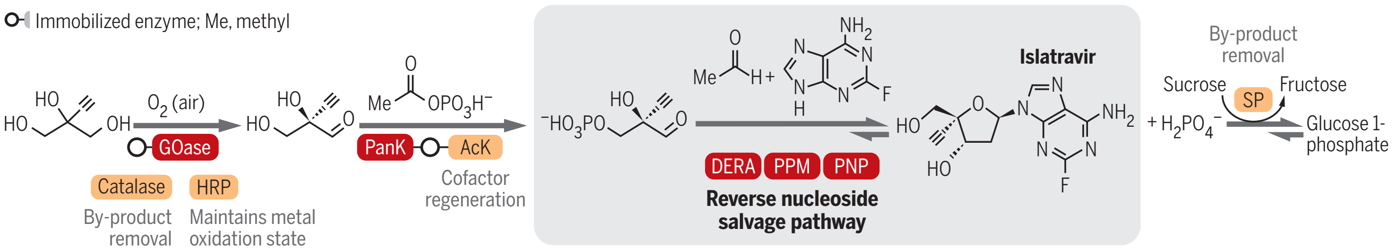
One of the many advantages of performing biocatalytic cascade reactions is the effective displacement of unfavorable reaction equilibria that can be achieved through product removal. However, despite performing the PNP and PPM steps in tandem, the reaction proceeded with poor conversion, and the inorganic phosphate by-product inhibits the enzymes. An elegant solution to these issues was the inclusion of an auxiliary sucrose phosphorylase, along with its sugar substrate, which removed free phosphate and effectively displaced the reaction equilibrium toward product formation.

Having assembled enzymes for the three key steps in the cascade, Huffman *et al.* sought to develop a biocatalytic route for the synthesis of the DERA substrate 2-ethynylglyceraldehyde 3-phosphate. Extensive screening of a broad range of kinases resulted in the identification of pantothenate kinase (PanK) from *E. coli*, which displayed low levels of activity (~1% conversion) toward the (*R*)-enantiomer of the target aldehyde. Despite the modest initial activity, directed evolution was successfully used to substantially improve the productivity and stability of this enzyme. Finally, after 12 rounds of evolution, the authors reversed the enantioselectivity and improved the activity, stability, and expression of a galactose oxidase variant for the desymmetrization of the starting substrate, 2-ethynylglycerol.

Advancements in protein engineering,

Engineering a biocatalytic cascade

The synthetic pathway for the production of the investigational drug islatravir (MK-8591) was inspired by the natural bacterial nucleoside salvage pathway. The cascade created by Huffman *et al.* uses a total of nine enzymes, five of which were evolved to optimize their properties and four additional auxiliary enzymes.



EVOLVED ENZYME	ABBREVIATION	SOURCE OF ENZYME
Galactose oxidase	GOase	Fungal
Panthothenate kinase	PanK	Bacterial
Deoxyribose-5-phosphate aldolase	DERA	Bacterial
Phosphopentomutase	PPM	Bacterial
Purine nucleoside phosphorylase	PNP	Bacterial

AUXILIARY ENZYME	ABBREVIATION	SOURCE OF ENZYME
Catalase	–	Mammalian
Horseradish peroxidase	HRP	Plant
Acetate kinase	AcK	Bacterial
Sucrose phosphorylase	SP	Bacterial

GRAPHIC: A. KITTERMAN/*SCIENCE*

rapid gene sequencing, and the availability of low-cost DNA synthesis have made it possible to alter the properties of enzymes and fine-tune them for biocatalytic applications (*6*–*8*). The work by Huffman *et al.* is a milestone in cascade design, largely because of the number of biocatalysts operating in tandem and the engineering feat required to optimize five of the nine enzymes involved in the synthesis. It also highlights how biosynthetic or degradative pathways can be a source of inspiration for the design of efficient biocatalytic cascades and how sequences can be reconstituted using enzymes recruited from multiple sources—in this case, of bacterial, fungal, plant, and mammalian origin. The diverse role that biocatalysts can play is also exemplified in this work, where five engineered enzymes are directly involved in the synthesis of the target molecule, and four additional enzymes function to recycle coenzyme, remove inhibitory by-products, and maintain the correct oxidation state of the copper cofactor.

Previous approaches reported for the synthesis of islatravir relied on multistep syntheses and require protecting group manipulations and intermediate purification steps (*9*, *10*). The incorporation of a key biocatalytic step or steps has the potential to revolutionize synthetic design strategies by making possible transformations that are not accessible using solely chemical approaches (*11*, *12*). The application of enzymes in industry and the development of chemoenzymatic routes to complex molecules is now well established. However, multistep syntheses exclusively comprising biocatalytic transformations are rare (*13*), and this contribution sets a new standard for the synthesis of complex molecules with enzymatic cascades. ■

REFERENCES AND NOTES

1. S. P. France, L. J. Hepworth, N. J. Turner, S. L. Flitsch, *ACS Catal.* **7**, 710 (2017).
2. S. Gandomkar, A. Żądło-Dobrowolska, W. Kroutil, *ChemCatChem* **11**, 225 (2019).
3. P. Both *et al.*, *Angew. Chem. Int. Ed.* **55**, 1511 (2016).
4. M. A. Huffman *et al.*, *Science* **366**, 1255 (2019).
5. C. K. Savile *et al.*, *Science* **329**, 305 (2010).
6. F. H. Arnold, *Angew. Chem. Int. Ed.* **57**, 4143 (2018).
7. C. Zeymer, D. Hilvert, *Annu. Rev. Biochem.* **87**, 131 (2018).
8. C. A. Denard, H. Ren, H. Zhao, *Curr. Opin. Chem. Biol.* **25**, 55 (2015).
9. M. McLaughlin *et al.*, *Org. Lett.* **19**, 926 (2017).
10. M. Kageyama, T. Nagasawa, M. Yoshida, H. Ohrui, S. Kuwahara, *Org. Lett.* **13**, 5264 (2011).
11. N. J. Turner, E. O'Reilly, *Nat. Chem. Biol.* **9**, 285 (2013).
12. M. Hönig, P. Sondermann, N. J. Turner, E. M. Carreira, *Angew. Chem. Int. Ed.* **56**, 8942 (2017).
13. S. Wu *et al.*, *Nat. Commun.* **7**, 11917 (2016).

ACKNOWLEDGMENTS

J.R. acknowledges the School of Chemistry, University College Dublin, for support.

10.1126/science.aaz7376

Lava erupts from a rift zone fissure during the 2014–2015 Bárðarbunga caldera collapse in Iceland.

VOLCANOLOGY

Calderas collapse as magma flows into rifts

Recent caldera collapses show the importance of distant volcanic rift zones

By **Freysteinn Sigmundsson**

Major magma drainage from volcanoes causes the collapse of volcanic edifices, forming calderas that can be both many kilometers wide and hundreds of meters deep. Many calderas form during major explosive eruptions, when magma erupts from fractures on ring faults that bound calderas. However, the most recent caldera-forming events at Kīlauea Volcano, Hawai'i, in 2018 (*1*) and at Bárðarbunga, Iceland, in 2014–2015 (*2*) formed by a different mechanism. Both events were gradual caldera collapses that occurred as magma flowed over long distances into rifts far away from volcano summits. The caldera ring faults associated with collapse only began to move after a major magma withdrawal into a rift zone.

The detailed monitoring of the recent Kīlauea collapse reveals the behavior of the magma plumbing system involved and the dynamic processes related to coupling a caldera collapse with a rift eruption. Anderson *et al.* used modeling of geodetic changes to give an account of the pressure drop in a magma body before and during a collapse (*3*). Geochemical analysis of the eruptive products by Gansecki *et al.* reveal how pockets of magma of different types were captured by an advancing dike and erupted at the beginning of activity (*4*). Patrick *et al.* demonstrate how some of the short-term fluctuations in eruption activity, witnessed as magma surges, were a response to pressure variations under the collapsing caldera (*5*). When compared to observations from Bárðarbunga (*2*, *6*–*10*) and other collapses, a pattern emerges that may be a typical and common way for calderas to form at basaltic volcanoes.

The beginning of the process is a large-scale lateral subsurface injection of magma into a volcano rift zone, which is a preferred path of dikes that forms in response to imposed stresses. Rift zones are common on volcanoes and result from the tectonic setting, topography, slope instability, and/or, for ocean island volcanoes, seaward sliding of a volcano flank. In Iceland, the divergence of tectonic plates causes the formation of rift zones. At Kīlauea Volcano, the cause is the gradual seaward sliding of the south flank of

Nordic Volcanological Center, Institute of Earth Sciences, University of Iceland, Reykjavík, Iceland. Email: fs@hi.is

PHOTO: RAGNAR TH. SIGURDSSON/AGEFOTOSTOCK/NEWSCOM

the volcano. If the conditions are right when the magma body under a caldera fails, an unusually large amount of magma may flow into rifts (see the figure). One example of this is when tectonic stretching over a long time has built up sufficient stress. At Kīlauea, initial magma injection into the rift may have caused stress to increase, triggering a large earthquake, on 4 May 2018, on a subhorizontal basal fault under Kīlauea's south flank (*1*) that in turn caused additional stretching across the rift zone.

The amount of magma transported into a rift zone can induce slip on caldera ring faults if it is large enough. Anderson *et al.* reveal, in detail, the deformation of Kīlauea before the onset of slip on the caldera faults and conclude that calderas may begin to collapse after removal of only a small fraction of stored magma.

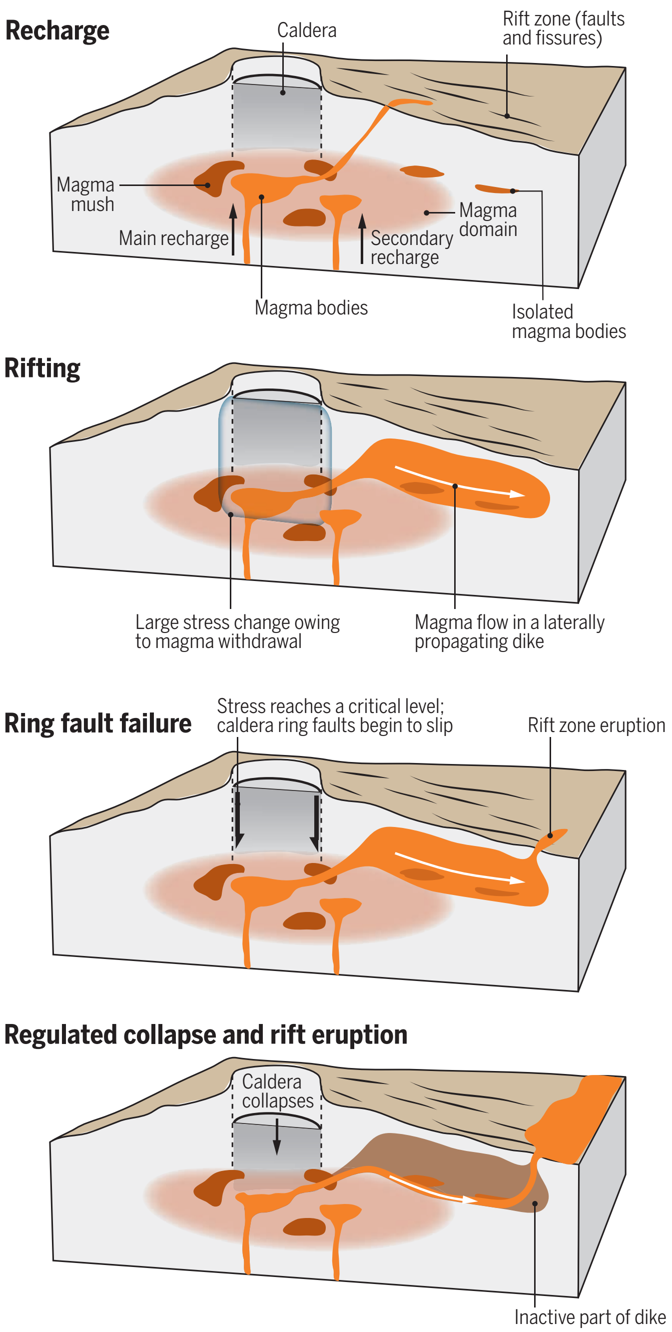
Forming new, active parts of the volcano plumbing system occurs by way of a laterally advancing dike in a rift zone. The dikes may hit crystal-rich layers called magma mush or pockets of magma residing either under the central part of the volcano or in the rift zone. These pockets of magma become part of the plumbing system as magma continues to advance. For Kīlauea, Gansecki *et al.* inferred that the rift zone had at least two separately stored magmas that the advancing dike intersected, which erupted first during the event. At Bárðarbunga, one magma type erupted, but crystals from crystal mush were incorporated into the magma (*11*, *12*). Both observations suggest an important role for these magma pockets in the chemistry of the dike magma and the timing for when different magma compositions are erupted.

After the plumbing system is stabilized as the rift eruption and central caldera collapse is ongoing, pressure governs the system evolution to a large extent. The central block within a caldera drives out a large volume of magma as it subsides. At Bárðarbunga, the mass flow rate and caldera collapse were found to decline almost exponentially throughout the eruption (*2*, *13*). Modeling indicated that the effective cross-sectional area of the flow path during the eruption (*2*) was only a small fraction of the cross-sectional area of the previously established dike (*6*). At Kīlauea, magma surges at the eruption site were linked to steps in the down-sagging of the caldera. At Bárðarbunga, such steps also occurred and influenced earthquake activity in the dike. These effects are similar to the water-hammer effects in fluid-filled pipes, resulting from sudden changes in pressure and hydraulic transients that travel in pipes (*14*). The observations show that variable pressure under a collapsing central block translates directly into eruption activity.

The recent observations of collapses have serious implications for hazards and volcano monitoring. Eruptions in rifts require monitoring of conditions at calderas far away from eruption activity that may provide crucial information on eruptions. This issue is more relevant at basaltic calderas than at other volcanoes because basalt is less viscous and flows more easily over long distances than other magma types. Although the next caldera formation may not have a large flank eruption, the recent detailed monitoring demonstrates that magma under collapsing calderas is under pressure, and copious amounts of magma can be driven toward far-away effusive eruptions. This sort of long-distance volcanic plumbing system appears to be a far more common occurrence than previously believed. ■

Caldera collapse coupled to rift zone eruption

Progressive steps in caldera formation in relation to lateral transfer of magma into rifts (generalized cross-section along a volcano rift). Recharge into the magma domain (*15*)—a complex of liquid magma bodies and crystal mush—occurs before an eruption. Rifting is followed by ring-fault failure and finally caldera collapse at the center of the volcano, regulated by a major eruption in the rift.



REFERENCES AND NOTES

1. C. A. Neal *et al.*, *Science* **363**, 367 (2019).
2. M. T. Gudmundsson *et al.*, *Science* **353**, 1 (2016).
3. K. R. Andersen *et al.*, *Science* **366**, eaaz1822 (2019).
4. C. Gansecki *et al.*, *Science* **366**, eaaz0147 (2019).
5. M. R. Patrick *et al.*, *Science* **366**, eaay9070 (2019).
6. F. Sigmundsson *et al.*, *Nature* **517**, 191 (2015).
7. G. Pedersen *et al.*, *J. Volcanol. Geotherm. Res.* **340**, 155 (2017).
8. T. Ágústsdóttir *et al.*, *J. Geophys. Res.* **124**, 8331 (2019).
9. J. Woods *et al.*, *Earth Planet. Sci. Lett.* **490**, 216 (2018).
10. M. M. Parks *et al.*, *Earth Planet. Sci. Lett.* **262**, 212 (2017).
11. M. E. Hartley *et al.*, *Contrib. Mineral. Petrol.* **173**, 10 (2018).
12. S. A. Halldórsson *et al.*, *Contrib. Mineral. Petrol.* **173**, 64 (2018).
13. D. Coppola *et al.*, *Geology* **45**, 523 (2017).
14. B. S. Jung, B. Karney, *Urban Water J.* **14**, 502 (2016).
15. F. Sigmundsson, *Geophys. Res. Lett.* **43**, 12423 (2016).

ACKNOWLEDGMENTS

Support comes from the Horizon 2020 project EUROVOLC, funded by the European Commission (grant 731070).

10.1126/science.aaz7126

POLICY FORUM

BIOMEDICAL TECHNOLOGY REGULATION

Algorithms on regulatory lockdown in medicine

Prioritize risk monitoring to address the "update problem"

By **Boris Babic[1], Sara Gerke[2], Theodoros Evgeniou[1], I. Glenn Cohen[3]**

As use of artificial intelligence and machine learning (AI/ML) in medicine continues to grow, regulators face a fundamental problem: After evaluating a medical AI/ML technology and deeming it safe and effective, should the regulator limit its authorization to market only the version of the algorithm that was submitted, or permit marketing of an algorithm that can learn and adapt to new conditions? For drugs and ordinary medical devices, this problem typically does not arise. But it is this capability to continuously evolve that underlies much of the potential benefit of AI/ML. We address this "update problem" and the treatment of "locked" versus "adaptive" algorithms by building on two proposals suggested earlier this year by one prominent regulatory body, the U.S. Food and Drug Administration (FDA) (*1*, *2*), which may play an influential role in how other countries shape their associated regulatory architecture. The emphasis of regulators needs to be on whether AI/ML is overall reliable as applied to new data and on treating similar patients similarly. We describe several features that are specific to and ubiquitous in AI/ML systems and are closely tied to their reliability. To manage the risks associated with these features, regulators should focus particularly on continuous monitoring and risk assessment, and less on articulating ex-ante plans for future algorithm changes.

THE REGULATORY DESIGN PROBLEM

One of the key advantages of AI/ML is that it can enable a "learning healthcare system," wherein the boundaries between research and practice are regarded as porous (*3*). Once the AI/ML is deployed, it can learn and thereby alter its performance and behavior, much the way a medical resident learns on the job. But this poses a difficult regulatory design challenge. Consider two polar approaches to the update problem:

One pole would be for a regulator to permit marketing of only a locked algorithm and require any change to the algorithm to undergo a completely new premarket review. Such an approach has several drawbacks. Suppose an algorithm for analyzing results of mammograms and making recommendations on breast cancer risk receives marketing authorization (*4*). Suppose the training data were underinclusive of African-American women who tend to have differences in breast density from Caucasian women. The algorithm would thus produce recommendations ill-suited for that population. As the AI/ML system is used in clinical settings that include more African-American women, it becomes possible to more accurately estimate the parameters used to predict breast cancer in this subpopulation when making recommendations.

Although improvements in pre-specified testing of subgroups might provide some benefit in avoiding this problem, in some situations, relevant subpopulations may not be known ex-ante. For example, in conducting HIV vaccine studies, researchers did not (and perhaps could not) know ex-ante that in a particular trial, the vaccine might increase rather than reduce HIV infection risk for "uncircumcised men who both had sex with men (MSM) and had high titers of preexisting antibodies against Ad5" (*5*). Prespecified testing is unlikely to capture these kinds of issues. Going forward, AI/ML may be able to identify such subpopulations and even develop customized models for different ones (some of which are only possible to identify after using AI/ML on lots of patients). Such customization would be health-promoting, but if another premarket review is needed, the update may never occur—the maker may not have a financial incentive to pursue the cost of another review and might also worry about the message it might send about the quality of its existing algorithm.

The opposite pole would be to treat the initial marketing authorization as permitting the AI/ML maker to update the algorithm without any further regulatory review. Such updates can be either of the algorithm itself ("algorithm updates")—for example, replacing a linear ML model with a polynomial one—or of the algorithm's parameters ("parametric updates"), which may be continuously tuned as the system is applied to new data in practice. This approach is likewise perilous. Parametric updates are at the core of modern AI/ML systems—they take place almost continuously, without human

[1]INSEAD, Singapore and Fontainebleau, France. [2]The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School; The Project on Precision Medicine, Artificial Intelligence, and the Law, Cambridge, MA, USA. [3]Harvard Law School, Cambridge, MA, USA. Email: igcohen@law.harvard.edu

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

Published by AAAS

input, and their effects can be hard to identify ex-ante. But the quality of parametric updates depends largely on the quality of the associated underlying data. An adaptive system with continuously changing parameters is susceptible to data quality issues that can arise from, for example, errors of the AI/ML users or intentional adversarial attacks (*6*). The latter can take many forms. Consider a hypothetical example (*6*): In response to the opioid crisis, many insurers now use patient- or provider-level overdose risk-prediction algorithms to deny oxycontin prescriptions. A physician, certain that a patient is in need of a prescription, may learn that the patient can avoid the algorithmic gatekeeper and secure a prescription by typing in a combination of codes that will guarantee a low risk for overdose score. Such a system incentivizes the elicitation of low-quality physician data. An unchecked dynamic algorithm would inappropriately adapt to this over time—considering all outcomes of prescriptions—and begin to falsely categorize low-risk patients as high risk. In this kind of situation, continuous oversight can provide a necessary check on adaptive AI/ML systems.

In an attempt to steer between these two poles, the FDA released a discussion paper in April 2019 (*1*). Until now, the FDA has exclusively approved or cleared medical AI/ML-based software as a medical device—what the FDA calls "SaMD," which is software that is on its own a medical device and is not part of a hardware medical device (*7*)—with "locked" algorithms (*1*). A locked algorithm is defined by the FDA as "an algorithm that provides the same result each time the same input is applied to it and does not change with use" (*1*). Any AI/ML system can satisfy this definition provided it is fixed in advance.

However, most AI/ML algorithms are "adaptive," arguably their key strength. Even parameters in a simple model like a logistic regression will gradually evolve as the model is refit in response to new data. For adaptive AI/ML-based SaMD, the FDA proposed a "total product lifecycle (TPLC) regulatory approach" that permits continuous improvement of such devices while maintaining their safety and effectiveness (*1*). The FDA's TPLC approach is a feature of the Software Precertification (Pre-Cert) Program that it is piloting on a small number of companies to determine its feasibility (*2*). One major idea in the FDA's April 2019 discussion paper is that AI/ML-based SaMD could be updated to a certain extent after marketing authorization; when seeking initial premarket review of an AI/ML-based SaMD, manufacturers would be given the option to submit a "predetermined change control plan," which would contain a description of anticipated modifications and an "Algorithm Change Protocol," including the associated methodology being used to implement such changes (*1*).

UNDERSTANDING RISKS

Before considering adaptive algorithms, it is important to recognize that a "locked" algorithm, as defined, for example, by the FDA, could be more harmful than an "adaptive" one—and vice versa.

To begin with, the concept of "locked" is ambiguous. We focus on two definitions that we call "system lock" and "true function lock." Do we want the AI/ML system to continually use the locked estimate of the function, relating inputs and outputs, that was first approved? This is how the FDA has defined what it means to be "locked," a concept that we call "system lock." Merely achieving "system lock" will not guarantee that the system is safe for patients. An alternative, and perhaps a preferable one, is that the algorithm locks, as closely as possible, to the true function that relates the inputs and output—which is unknown ex-ante in practice and which emerges over time. We call this "true function lock."

For adaptive algorithms, it is especially important for regulators to assess whether the AI/ML system is overall reliable as applied to new data—i.e., whether it approaches "true function lock." Several AI/ML features are identified below that, when not properly considered, can lead the AI/ML system to use a poor estimate of the true relationship between the inputs and outputs and thereby possibly cause harm to patients (for example, through misdiagnosis). Regulators need to focus their attention on such issues in order to manage the risk that AI/ML systems learn and use a wrong input-output relation.

Concept drift

Concept drift describes a situation where the true relation between inputs and outputs changes (over time). This may happen because of a changing environment or because the model was misspecified (for example, the estimated function is linear when the actual relationship is quadratic, or there are omitted variables, etc.). Consider, for example, an AI/ML system trained to identify skin lesions as benign or malignant (*8*). The model presupposes an underlying distribution of these labels (benign versus malignant). However, the datasets that these AI/ML systems rely on typically do not track race or skin color, or may miss or not report certain skin types. Yet the malignancy of skin lesions (the true relation between input and output/diagnosis) may vary across race and skin type. As a result, the same image can lead to two different probabilistic diagnoses, depending on the underlying skin/race, an omitted feature. This sort of problem is ubiquitous in medical AI/ML.

A regulatory regime requiring a "system lock" is not immune to this problem. Indeed, a "system locked" algorithm can make matters worse by prohibiting the system from learning. Moreover, a regime focused on predetermined change control plans is likewise vulnerable to risks arising from concept drift. Any predetermined change control plan risks being either uninformative or impractical—depending on the level of detail at which a maker would be expected to describe future modifications. At one extreme, a maker might be required to describe proposed changes in very general terms. This would be uninformative. On the other extreme, they might be required to describe precisely the sorts of changes they anticipate. Such a task is not feasible without having seen all possible future data from all types of patients and conditions—especially for AI/ML algorithms that may have thousands or millions of parameters. Even if (in theory, or possibly with future technologies) this kind of task could be accomplished, it would be extremely difficult and time-consuming—and thus, impractical. Moreover, such a plan could be especially harmful when unanticipated problems are reported—in which case, the proposed framework could require another round of review.

Covariate shift

When the input distribution of new data is different from the data that the algorithm was trained or tested for approval on, the result is covariate shift (*9*). This can occur in the absence of concept drift, although the two are not mutually exclusive. For example, training data may have come from geographically centralized clinical sites, but the device is to be deployed beyond those regions and populations. When this occurs, "system locking" the algorithm hampers the maker's ability to address the problem. Further, describing how the distribution of patients may change is not something a maker may be able to do ex-ante because they usually do not know the distribution of the data that the algorithm will be applied to.

Instability

One major concern is treating similar patients similarly. That is, medically insignificant differences among patients should not lead to substantive differences in diagnosis or treatment. Suppose that when an AI/ML system is given a set of inputs, it produces one probabilistic output. For example, the probability that a particular skin lesion is malignant is 87%. Now suppose that very small changes are made to the set of inputs provided to the underlying algorithm. For ex-

ample, the values associated with a few pixels may be changed in a way that is medically insignificant. The AI/ML system must now make a prediction from a feature vector that is vanishingly different from the initial vector. A stable algorithm should give predictions that are similarly "close" in the output space (in probability) when it is given a slight variation relative to another input [Dwork *et al.* (*10*) describe this property and use it as the basis for a definition of fairness in ML].

When this condition is not satisfied, the algorithm is not stable in the sense that medically similar patients can receive dissimilar diagnoses. From the perspective of patient safety, it is undesirable to have a diagnostic system that frequently classifies medically similar lesions very differently. Paying attention to this encourages thinking not in terms of same inputs/same outputs, but in terms of similar inputs/similar outputs.

In modern AI/ML, leading classification systems are highly nonlinear. This makes them especially vulnerable to such instability [for example, (*11*)]. This problem extends beyond adversarial attacks (*6*). "System locking" of an algorithm, though it can guarantee that the same inputs will lead to the same outputs, does not secure against the bigger concern—instability. Meanwhile, any predetermined change control plan does not get to the core of the problem either, because it is impossible to know in advance what kind of instabilities the world actually has.

A CONTINUOUS RISK-MONITORING APPROACH

As regulators push forward, their emphasis should be on developing a process to continuously monitor, identify, and manage associated risks due to AI/ML features such as concept drift, covariate shift, and instability. Such a process can include, for example, the following elements, some of which might even be automated with improvements in AI/ML technology:

Retesting

An AI/ML system may need to be regularly retested (possibly continuously retested, with dedicated infrastructure) on all past cases, or a random subset of them, including but not limited to the ones used for the initial marketing authorization. Major discrepancies on past verdicts may lead to regulatory action.

Simulated checks

An AI/ML system can be continuously applied to "simulated patients"—generated, for example, by perturbing the data of past patients, an idea often used to examine the robustness of AI/ML models—to evaluate whether its behavior is reliable with respect to a sufficient diversity of patient types.

Adversarial stress tests

Every AI/ML system may need to be paired with a monitoring mechanism to ensure robustness to adversarial examples (*12*). Regulators could use the adversarial approach to conduct algorithmic stress tests throughout the AI/ML system's lifecycle, borrowing practices such as red teaming and adversarial attack testing from cybersecurity.

An appropriate division of labor

Monitoring of AI/ML systems should, in general, be done by actors different from the ones developing these systems. Separation of development and testing is common in other contexts. For example, in software development, quality assurance and development teams are separate, while risk management and compliance departments are separated from traders in the financial sector. Such divisions may be likewise required for companies developing medical AI/ML systems. Moreover, third-party organizations that monitor AI/ML systems based on standards the industry develops, similar in spirit to those of professional organizations like the Institute of Electrical and Electronics Engineers and the International Organization for Standardization, may also play a role in the future.

Use of innovative electronic systems

Regulators could also use new electronic systems and data analysis techniques, such as change-point detection (a family of statistical techniques that attempt to identify changes in the distribution of a stochastic process) or anomaly detection (a family of techniques used to identify rare items in a data set), to continuously monitor AI/ML systems. Existing elements of regulators' product oversight could be adapted for monitoring AI/ML. For example, systems such as the FDA's national medical product monitoring system Sentinel (*13*, *14*) could be used to continuously monitor the behavior of approved AI/ML-based medical devices. Combining information from electronic health records and other data from such devices, regulators could themselves perform some of the tasks described. Indeed, in September 2019, the FDA announced an intention to expand Sentinel to three separate coordinating centers monitoring more traditional medical product safety (*13*). Its new "Operations Center" seeks to use partnerships in epidemiology, statistics, and data science among other fields (*13*). Its new "Innovation Center" will explore new roads "to extract and structure information from electronic health records" (*13*). Both of these new centers could be used to implement aspects of the AI/ML monitoring proposed here..

Our suggestions refine the FDA's aim of implementing "real-world performance monitoring" (*1*, *2*) by articulating some key features that risk-monitoring should focus on (i.e., concept drift, covariate shift, and instability) and suggesting some ways to implement it. In principle, our goal is to emphasize the risks that can arise from unanticipated changes in how medical AI/ML systems react or adapt to their environments. Subtle, often unrecognized parametric updates or new types of data can cause large and costly mistakes. Hence, they need to be continuously monitored and tested. Although this discussion has been motivated by the FDA's current approach to AI/ML-based SaMD and the U.S. experience, with appropriate adaptations, the lessons here apply to other countries and their regulators as well. ■

REFERENCES AND NOTES

1. U.S. Food and Drug Administration (FDA), "Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)" (discussion paper and request for feedback, 2019); www.fda.gov/media/122535/download.
2. U.S. Food and Drug Administration (FDA), "Developing a Software Precertification Program: *A Working Model*" (v.1.0, January 2019); www.fda.gov/media/119722/download.
3. The learning healthcare project, "Background, learning healthcare system" (2019); www.learninghealthcareproject.org/section/background/learning-healthcare-system.
4. A. Yala, C. Lehman, T. Schuster, T. Portnoi, R. Barzilay, *Radiology* **292**, 60 (2019).
5. A. S. Fauci, M. A. Marovich, C. W. Dieffenbach, E. Hunter, S. P. Buchbinder, *Science* **344**, 49 (2014).
6. S. G. Finlayson *et al.*, *Science* **363**, 1287 (2019).
7. US Food and Drug Administration (FDA), "Software as a medical device (SaMD)" (2018); www.fda.gov/medical-devices/digital-health/software-medical-device-samd.
8. A. Esteva *et al.*, *Nature* **542**, 115 (2017).
9. S. Bickel, M. Bruckner, T. Scheffer, Discriminative learning for differing training and test distributions, Proceedings of the 24th International Conference on Machine Learning (ICML), Corvallis, OR, 2007.
10. C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel, Fairness through awareness, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS), Cambridge, MA, 2012.
11. I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, 2015.
12. S. Gu, L. Rigazio, Towards deep neural network architectures robust to adversarial examples, Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, 2015.
13. U.S. Food and Drug Administration (FDA), "FDA's Sentinel Initiative" (2019); www.fda.gov/safety/fdas-sentinel-initiative.
14. Sentinel Coordinating Center, "Sentinel is a national medical product monitoring system" (2019); www.sentinelinitiative.org.

ACKNOWLEDGMENTS

S.G. and I.G.C. were supported by a grant from the Collaborative Research Program for Biomedical Innovation Law, a scientifically independent collaborative research program supported by a Novo Nordisk Foundation grant (NNF17SA0027784). All authors contributed equally to the analysis and drafting of the paper. I.G.C. served as a bioethics consultant for Otsuka on their Abilify MyCite product. The authors declare no other competing interests.

10.1126/science.aay9547

LETTERS

Wide-scale adoption of wind energy poses social and conservation challenges.

Edited by **Jennifer Sills**

Wind energy: A human challenge

In their Review "Grand challenges in the science of wind energy" (25 October, p. eaau2027), P. Veers *et al.* outline an "integrative" vision centered on three legs they contend are "critical to realizing the full potential of wind energy": improving understanding of the physics of flow in and around wind power projects, engineering of large machines, and integration of wind into the electricity grid. However, large-scale wind power implementation is unlikely to succeed without the social sciences. Although wind power deployment on a massive scale is a technical question, it is even more so a public issue.

Persuading people to adopt new technology requires an understanding of the relationships people have with places and landscapes and an appreciation of how they perceive risk (*1–3*). Conflict resolution strategies will have to be put in place to moderate the competing uses of land and sea, such as offshore wind power and commercial fishers (*4*). Institutions and developers will have to build trust (*5*). Public participation must be encouraged to ensure that those affected are given a voice and some influence over the outcome (such as the number of turbines or the project layout) (*5*). The transition to renewable energy ought to be implemented in a way that considers those communities long reliant on fossil fuel development (*6*). The inequitable burden placed on those living near transmission lines must be addressed so that needed transmission can actually be built (*7*). Without a thorough understanding of the human context (*8–10*), society may fall well short of "realizing the full potential of wind energy."

Jeremy Firestone
College of Earth, Ocean, and Environment, University of Delaware, Newark, DE 19716, USA.
Email: jf@udel.edu

REFERENCES AND NOTES

1. P. Devine-Wright, *J. Commun. Appl. Soc. Psychol.* **19**, 426 (2009).
2. M. J. Pasqualetti, *Ann. Assoc. Am. Geogr.* **101**, 907 (2011).
3. B. Van Veelen, C. Haggett, *Sociol. Rural.* **57**, 533 (2016).
4. J. Rand, B. Hoen, *Energ. Res. Soc. Sci.* **29**, 135 (2017).
5. J. Firestone *et al.*, *J. Environ. Pol. Plan.* **20**, 370 (2018).
6. J. Firestone, H. Kirk, *Nat. Energ.* **4**, 311 (2019).
7. J. N. Cain, H. T. Nelson, *Land Use Pol.* **33**, 204 (2013).
8. D. Bell, T. Gray, C. Haggett, *Environ. Polit.* **14**, 460 (2005).
9. R. Wüstenhagen, M. Wolsink, M. J. Bürer, *Energ. Pol.* **35**, 2683 (2007).
10. M. Wolsink, *Energ. Res. Soc. Sci.* **46**, 287 (2018).

COMPETING INTERESTS

J.F. for the University of Delaware serves as managing director of a private joint venture between the University and Siemens Gamesa that owns and operates a 2MW wind turbine that supplies energy to the campus and town of Lewes, Delaware, and uses revenues to fund research and student fellowships. He also consults for Lawrence Berkeley National Laboratory.

10.1126/science.aaz8932

Wind energy: An ecological challenge

In their Review "Grand challenges in the science of wind energy" (25 October, p. eaau2027), P. Veers *et al.* identify three grand challenges in the science of wind energy based in atmospheric physics, material science, and electrical engineering. However, other crucially important challenges include understanding the effects of wind energy production on ecological systems and developing and deploying tools to mitigate negative environmental effects.

Wind energy production affects species and ecological systems in three ways. Collisions with wind turbines kill volant organisms, such as birds and bats, sometimes causing population-level consequences (*1*, *2*). Wind facilities and associated infrastructure transform the landscape, contributing to habitat loss, degradation, and fragmentation (*3*, *4*) and altering species behavior (*5*, *6*). Finally, wind turbines alter microscale and possibly macroscale weather (*7–9*).

Globally, environmental issues cost wind developers and operators millions of dollars and halt construction of facilities [e.g., (*10*, *11*)]. Wildlife fatalities and habitat loss are the foundation of legal prosecutions [e.g., (*10*, *11*)]. The build-out goals outlined by Veers *et al.* require an increase in the number of turbines and expansion into new sites, potentially exacerbating these problems. Moreover, the engineering-related developments required to meet Veers *et al.*'s grand challenges will result in taller, larger wind turbines, with longer, faster-moving blades that produce more energy but may also increase adverse effects on the environment.

Engineers, atmospheric scientists, and industry representatives are working collaboratively with conservation scientists to develop research, applied science, best practices, and management actions aimed at reducing the environmental impacts of wind energy. Efforts include modeling wildlife movements, fatalities, and risk to individuals and populations; developing technologies to deter bats and birds from wind turbines;

PHOTO: ISTOCK.COM/GABRIEL SHAKOUR

and deploying automated shutdown technologies that stop turbines in the presence of target species (*12*). Fortunately, many of the integrated and cross-disciplinary research efforts in atmospheric, physical, computational, and data science that Veers *et al.* identify contribute to the collaborative science required to address the environmental challenges as well.

Todd E. Katzner,[1] David M. Nelson[2]*, Jay E. Diffendorfer,[3] Adam E. Duerr[4,5,6], Caitlin J. Campbell[7], Douglas Leslie[8], Hannah B. Vander Zanden[7], Julie L. Yee,[9] Maitreyi Sur[10], Manuela M.P. Huso,[11] Melissa A. Braham[2,12], Michael L. Morrison[13], Scott R. Loss[14], Sharon A. Poessel,[1] Tara J. Conkling,[1] Tricia A. Miller[4,5]

[1]Forest and Rangeland Ecosystem Science Center, U.S. Geological Survey, Boise, ID 83706, USA. [2]Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD 21532, USA. [3]Geosciences and Environmental Change Science Center, U.S. Geological Survey, Denver, CO 80225, USA. [4]Conservation Science Global, West Cape May, NJ 08204, USA. [5]Division of Forestry & Natural Resources, West Virginia University, Morgantown, WV 26506, USA. [6]Bloom Research Inc., Santa Ana, CA 92705, USA. [7]Department of Biology, University of Florida, Gainesville, FL 32611, USA. [8]ICF International, Sacramento, CA 95814, USA. [9]Western Ecological Research Center, U.S. Geological Survey, Santa Cruz, CA 95060, USA. [10]Conservation Science Global, Shillong, Meghalaya, India. [11]Forest and Rangeland Ecosystem Science Center, U.S. Geological Survey, Corvallis, OR 97330, USA. [12]Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA. [13]Department of Wildlife and Fisheries Sciences, Texas A&M University, TX 77843, USA. [14]Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078, USA.
*Corresponding author. Email: dnelson@umces.edu

REFERENCES AND NOTES

1. W. F. Frick *et al.*, *Biol. Conserv.* **209**, 172 (2017).
2. T. E. Katzner *et al.*, *Conserv. Biol.* **31**, 406 (2017).
3. K. Barré *et al.*, *Biol. Conserv.* **226**, 205 (2018).
4. A. T. Marques *et al.*, *J. Anim. Ecol.* 10.1111/1365-2656.12961 (2019).
5. R. Łopucki, D. Klich, S. Gielarek, *Environ. Monit. Assess.* **189**, 343 (2017).
6. D. Fernández-Bellon *et al.*, *Conserv. Biol.* **33**, 413 (2019).
7. S. B. Roy, J. J. Traiteur, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17899 (2010).
8. L. M. Slawsky *et al.*, *Sensors* **15**, 14981 (2015).
9. J. K. Lundquist, K. K. DuVivier, D. Kaffine, J. M. Tomaszewski, *Nat. Energ.* **4**, 26 (2019).
10. R. Horsch *et al.*, "Court halts construction and limits operation of wind project for failure to comply with Endangered Species Act" (2009); www.lexology.com/library/detail.aspx?g=8999cf47-9ffe-478a-837d-25c45280fd1e.
11. U.S. Department of Justice, Office of Public Affairs, "Utility company sentenced in Wyoming for killing protected birds at wind projects" (2014); www.justice.gov/opa/pr/utility-company-sentenced-wyoming-killing-protected-birds-wind-projects-0.
12. T. D. Allison *et al.*, *Issues Ecol.* **21**, 1 (2019).

COMPETING INTERESTS

D.M.N. is a science adviser for the American Wind Wildlife Institute. A.E.D. has been paid as a consultant by wind energy companies. A.E.D., M.A.B., and T.A.M. have received research funding from wind energy companies and the American Wind Wildlife Institute. T.A.M. has also provided paid expert testimony in a hearing regarding golden eagle use of a wind energy facility in New York State. T.J.C. is a member of the Alameda County (California) Wind Repowering/Avian Protection Technical Advisory Committee.

10.1126/science.aaz9989

Protecting Patagonian peatlands in Chile

In their Letter "Seeing Chile's forest for the tree plantations" (27 September, p. 1388), A. P. Durán and O. Barbosa explain how Chile's current proposal for reducing greenhouse gas emissions (*1*) inadequately addresses forest management with exotic tree plantations. We agree, but we are even more concerned that the proposal overlooks other ecosystems entirely. Chilean Patagonian peatlands cover 3.1 million hectares (*2*) and contain approximately 4800 million tons of carbon accumulated over 18,000 years (*3*, *4*). This is 4.7 times more carbon than the aboveground biomass of forests in Chile (*4*, *5*). Peat in Chile is classified as a fossil resource, allowing it to be exploited by the Ministry of Mining (*6*). Chile should invest in the protection of this important ecosystem.

Because of the slow peat accumulation in sub-Antarctic regions (less than 1 mm per year) (*4*), exploitation of peatlands compromises their carbon sequestration capacity, shifting peatlands from net carbon sinks into net carbon sources (*7*). Protecting Chile's Patagonian peatlands would help the country achieve carbon neutrality by 2050 (*8*, *9*). To protect the peatlands, Chile must end their classification as fossil resources. Instead, Chile should present peatland preservation as part of its greenhouse gas reduction contributions at the 2019 United Nations Climate Change Conference (COP25) (now planned for Madrid, Spain, instead of Chile). Patagonian peatlands should also be recognized as overlooked carbon sinks of regional importance in Chile's new Climate Change Law (*10*).

Jorge Hoyos-Santillan[1,2,3]*, Alejandro Miranda[1,3,4], Antonio Lara[1,5,6], Maisa Rojas[1,7], Armando Sepulveda-Jauregui[1,3]

[1]Center for Climate and Resilience Research (CR)[2], Santiago, Metropolitan Region, Chile. [2]School of Biosciences, University of Nottingham, Sutton Bonington, Loughborough, UK. [3]Network for Extreme Environments Research, University of Magallanes, Punta Arenas, Magallanes, Chile. [4]Laboratorio de Ecología del Paisaje y Conservación, Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco, Cautín, Chile. [5]Instituto de Conservación, Biodiversidad y Territorio, Universidad Austral de Chile, Valdivia, Valdivia, Chile. [6]Fundación Centro de los Bosques Nativos—FORECOS, Valdivia, Valdivia, Chile. [7]Geophysics Department, University of Chile, Santiago, Metropolitan Region, Chile.
*Corresponding author. Email: jorge.hoyos@umag.cl

REFERENCES AND NOTES

1. Gobierno de Chile, "Contribución Determinada a Nivel Nacional (NDC) de Chile—Primera Actualización" (Santiago, Chile, 2019); https://mma.gob.cl/wp-content/uploads/2019/10/Propuesta_actualizacion_NDC_Chile_2019.pdf [in Spanish].

2. CONAF, CONAMA, "Monitoreo y actualización: Catastro de uso del suelo y vegetación, Región de Magallanes y Antártica Chilena" (Santiago, Chile, 2006); https://sit.conaf.cl/tmp/obj_632948/153_actualizacion_Magallanes.pdf [in Spanish].
3. V. Markgraf, U. M. Huber, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **297**, 351 (2010).
4. J. Loisel, in *Funciones y servicios ecosistémicos de las turberas en Magallanes-Colección de libros INIA N° 33*, E. Domínguez, D. Vega-Valdés, Eds. (Instituto de Investigaciones Agropecuarias, Punta Arenas, Chile, ed. 1, 2015), chap. 11 [in Spanish].
5. S. S. Saatchi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9899 (2011).
6. D. Villela, E. Castillo, J. Cantallopts, *Potencial de la turba en la minería no metálica en Chile* (Ministerio de Mineria, Santiago, Chile, 2017) [in Spanish].
7. J. Leifeld, L. Menichetti, *Nat. Commun.* **9**, 1071 (2018).
8. M. Maynard, "U.N. talks host Chile ramps up climate ambition—but is it enough?" *Reuters* (2019).
9. Intergovernmental Panel on Climate Change, "Climate Change and Land" (2019); www.ipcc.ch/site/assets/uploads/2019/08/4.-SPM_Approved_Microsite_FINAL.pdf.
10. Gobierno de Chile, "Anteproyecto de Ley Marco de Cambio Climático" (Santiago, Chile, 2019); https://mma.gob.cl/wp-content/uploads/2019/07/Presentacion-Ley-Marco-CC.pdf [in Spanish].

10.1126/science.aaz9244

TECHNICAL COMMENT ABSTRACTS

Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions"

S. Hesselmann, T. C. Lang, M. Schuler, S. Wessel, A. M. Läuchli

Tang *et al*. (Research Articles, 10 August 2018, p. 570) report on the properties of Dirac fermions with both on-site and Coulomb interactions. The substantial decrease, up to ~40%, of the Fermi velocity of Dirac fermions with on-site interaction is inconsistent with the numerical data near the Gross-Neveu quantum critical point. This results from an inappropriate finite-size extrapolation.

Full text: dx.doi.org/10.1126/science.aav6869

Response to Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions"

Ho-Kin Tang, J. N. Leaw, J. N. B. Rodrigues, I. F. Herbut, P. Sengupta, F. F. Assaad, S. Adam

Hesselmann *et al*. question one of our conclusions: the suppression of Fermi velocity at the Gross-Neveu critical point for the specific case of vanishing long-range interactions and at zero energy. The possibility they raise could occur in any finite-size extrapolation of numerical data. Although we cannot definitively rule out this possibility, we provide mathematical bounds on its likelihood.

Full text: dx.doi.org/10.1126/science.aav8877

EDITOR'S NOTE

"Joint statement on EPA proposed rule and public availability of data (2019)" by H. Thorp, M. Skipper, V. Kiermer, M. Berenbaum, D. Sweet, R. Horton, *Science* **366**, eaba3197 (2019). Published online 6 December 2019 (First Release 26 November 2019); 10.1126/science.aba3197

Science

Cite as: S. Hesselmann *et al.*, *Science*
10.1126/science.aav6869 (2019).

Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions"

S. Hesselmann[1], T. C. Lang[2], M. Schuler[2], S. Wessel[1], A. M. Läuchli[2]*

[1]Institute for Theoretical Solid State Physics, JARA-FIT and JARA-HPC, RWTH Aachen University, 52056 Aachen, Germany. [2]Institute for Theoretical Physics, University of Innsbruck, 6020 Innsbruck, Austria.

*Corresponding author. Email: andreas.laeuchli@uibk.ac.at

Tang *et al.* (Research Articles, 10 August 2018, p. 570) report on the properties of Dirac fermions with both on-site and Coulomb interactions. The substantial decrease, up to ~40%, of the Fermi velocity of Dirac fermions with on-site interaction is inconsistent with the numerical data near the Gross-Neveu quantum critical point. This results from an inappropriate finite-size extrapolation.

The low-energy excitations of many condensed matter systems, such as electrons on the honeycomb lattice of graphene, can be described by massless Dirac fermions with a Dirac cone-like dispersion relation and a corresponding Fermi velocity. The inclusion of interactions among the fermions eventually leads to a breakdown of this description, once the system undergoes a quantum phase transition to an insulating phase beyond a critical interaction strength. Below this interaction-induced quantum critical point (QCP), the system is characterized by massless Dirac fermions with a renormalized Fermi velocity. The quantification of this velocity renormalization constitutes a challenge in numerical simulations: Crossover effects strongly alter finite-size system estimates close to critical points, and a careful analysis of the actual excitation energies is required to extract reliable results.

Tang *et al.* (*1*) extract the momentum-resolved one-particle excitation energies from imaginary-time correlation functions obtained by projective quantum Monte Carlo (QMC) simulations. Upon approaching the Dirac points, the lattice dispersion of the noninteracting (tight-binding) fermion system takes on a linear, relativistic form that defines the tight-binding Fermi velocity v_0 at the Dirac point. The inclusion of either on-site (Hubbard) interactions or extended Coulomb interactions leads to changes of these excitation energies. Below the interaction-induced Gross-Neveu QCP, the dispersion remains gapless at the Dirac point in the thermodynamic limit (TDL) at infinite lattice size, defining the semimetallic (SM) regime. For the case of the Hubbard model, the Gross-Neveu QCP is known to be located at an on-site repulsion of $U_c(\gamma = 0) = 3.85(2)t$, beyond which the model exhibits antiferromagnetic order (*2*). Here, t denotes the nearest-neighbor hopping strength on the honeycomb lattice, and $\gamma = 3\alpha_0/U$ in terms of the Coulomb interaction strength α_0. Throughout this comment, we follow the notation used in (*1*).

In order to extract the interaction-induced renormalization of the Fermi velocity within the SM phase, the excitation gaps obtained from the QMC data for finite-size systems need to be extrapolated to the TDL. Finite-size effects are observed in all excitation energies, but in particular close to the QCP. This is seen in Fig. 1, which shows the bare finite-size excitation gaps, extracted from the imaginary-time QMC data as detailed in the supplementary materials of (*1*), based on the datasets made available online by the authors of (*1*). We observe that the finite-size effects are most pronounced at the Dirac points themselves (Fig. 1), where the gap vanishes in the TDL within the SM regime for $U < U_c(0)$ and at the Gross-Neveu QCP $U = U_c(0)$. On the other hand, for momenta in the immediate vicinity of the Dirac points, the finite-size effects are seen to be much weaker (Fig. 1), and one may estimate the TDL values of the excitation energies at these momenta from the values on the largest system sizes accessed in (*1*).

In Fig. 1 we also include data provided by Tang *et al.*, showing their finite-size extrapolated gaps. This processed data (based on the interpolation scheme used in their figure S2) are seen to be incompatible with the behavior of the excitation energies for small values of $a\Delta k$ extracted with our scheme. Moreover, as shown in Fig. 2A, the excitation energies close to, but excluding, the Dirac point exhibit only a weak dependence on U. Thus, for $\gamma = 0$, the low-energy Dirac dispersion, and hence the Fermi velocity, is in fact only weakly modified by the on-site interactions. In particular, the low-energy dispersion traced by our data in Fig. 1 for $U = 3.75t$ is clearly inconsistent with the ~40% decrease of the Fermi velocity from v_0 reported in (*1*), which is indicated by the lower red line in Fig. 1.

A reliable estimate for the Fermi velocity at the Dirac point for values of U inside the SM regime can be obtained from a finite-size analysis of the rescaled lowest particle-excitation energy $E/(a\Delta k)$ at the closest momentum to the Dirac point on each finite lattice. The corresponding finite-size values are compared to v_0 in Fig. 2B, and they demonstrate a remarkably weak renormalization of the Fermi velocity throughout the SM phase. A reduction by ~40% from the value v_0 is not compatible with the observed steady approach of $E/(a\Delta k)$ toward v_0 with increasing system size for all considered values of U within the SM regime.

The substantial overestimation of the Fermi velocity suppression by the on-site interaction reported in (*1*) (see also Fig. 2C) is in fact due to an inappropriate finite-size extrapolation procedure, which is documented in figure S2 of (*1*): The authors of (*1*) use the slope between the finite-size excitation energies at the Dirac point and the closest point to the Dirac point [with a linear interpolation to the simulation scale] as estimator. The finite-size energies at the Dirac point suffer from particularly large finite-size effects near the Gross-Neveu QCP, and the strong suppression of the Fermi velocity that is reported in (*1*) near the Gross-Neveu QCP merely reflects the enhanced finite-size effects of the excitation energy at the Dirac point, but not the renormalization of the actual low-energy dispersion. The extraction of velocities based on the softest excitations is also reported to be subtle for related quantum phase transitions [see, e.g., (*3*–*5*)].

Their means of data analysis therefore did not allow the authors of (*1*) to faithfully reproduce the Fermi velocity renormalization beyond the weak-coupling regime. The Fermi velocity renormalization shown in figure 2 of (*1*) is affected strongly by their finite-size analysis scheme, in particular in the vicinity of the Gross-Neveu QCP at $U_c(\gamma)$, which calls for a revised analysis and interpretation of the numerical data along the lines outlined in this comment.

REFERENCES

1. H. K. Tang, J. N. Leaw, J. N. B. Rodrigues, I. F. Herbut, P. Sengupta, F. F. Assaad, S. Adam, The role of electron-electron interactions in two-dimensional Dirac fermions. *Science* **361**, 570–574 (2018). doi:10.1126/science.aao2934 Medline
2. Y. Otsuka, S. Yunoki, S. Sorella, Universal quantum criticality in the metal-insulator transition of two-dimensional interacting Dirac electrons. *Phys. Rev. X* **6**, 011029 (2016). doi:10.1103/PhysRevX.6.011029
3. A. Sen, H. Suwa, A. Sandvik, Velocity of excitations in ordered, disordered, and critical antiferromagnets. *Phys. Rev. B* **92**, 195145 (2015). doi:10.1103/PhysRevB.92.195145
4. M. Schuler, S. Whitsitt, L.-P. Henry, S. Sachdev, A. M. Läuchli, Universal Signatures of Quantum Critical Points from Finite-Size Torus Spectra: A Window into the Operator Content of Higher-Dimensional Conformal Field Theories. *Phys. Rev. Lett.* **117**, 210401 (2016). doi:10.1103/PhysRevLett.117.210401 Medline
5. M. Schuler, S. Hesselmann, S. Whitsitt, T. C. Lang, S. Wessel, A. M. Läuchli, Torus Spectroscopy of the Gross-Neveu-Yukawa Quantum Field Theory: Free Dirac versus Chiral Ising Fixed Point. arXiv 1907.05373 [cond-mat.str-el] (11 July 2019).
6. S. Hesselmann *et al*., Replication data for: Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions"; https://doi.org/10.7910/DVN/SNQHQN.

ACKNOWLEDGMENTS

We thank H.-K. Tang and colleagues for making their data openly available. **Funding:** Supported by FWF projects I-2868-N27 and F4018 and by DFG projects RTG 1995 and FOR 1807. **Author contributions:** S.H., T.C.L., and M.S. performed the data analyses and prepared the figures; S.W. and A.M.L. directed the investigation; the manuscript reflects the contributions of all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and computer scripts are available at Harvard Dataverse (*6*).

24 October 2018; accepted 4 November 2019
Published online 6 December 2019
10.1126/science.aav6869

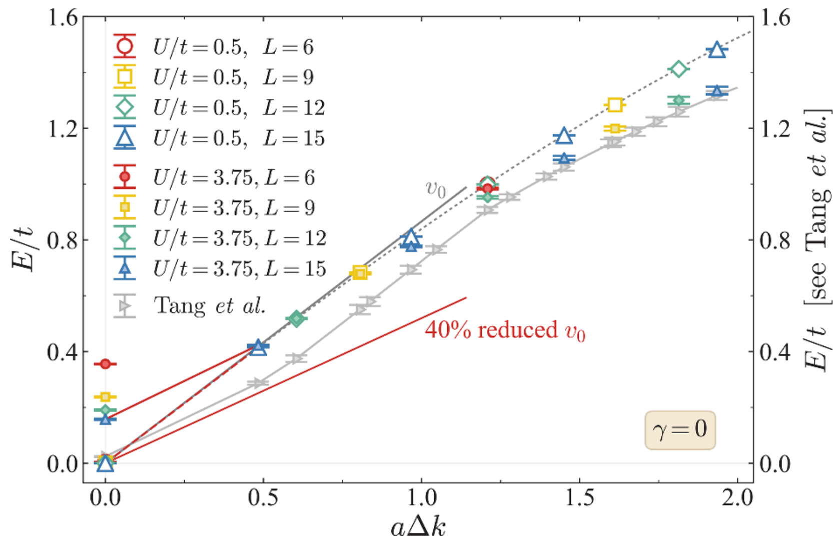


Fig. 1. Low-energy dispersions for the Hubbard model on the honeycomb lattice at different interaction strengths. Dependence of the bare lowest particle-excitation energy E on the distance $a\Delta k$ to the Dirac point is shown for the Hubbard model (γ = 0) on the honeycomb lattice at U/t = 0.5 and 3.75. E is deduced from the imaginary-time slope of the Green's function at the corresponding momenta for different linear lattice sizes L of the system. The dashed dark gray line traces the lattice dispersion relation for the tight-binding model of noninteracting fermions (U/t = 0). Also indicated are linear dispersions corresponding to v_0 (dark gray solid line) and to the 40% decrease with respect to v_0 reported in (*1*) (lower red solid line), and lines that connect the excitation energy at the Dirac point to its value at the nearest-neighbor momenta on the L = 15 lattice for U/t = 0.5 (dashed red line) and for U/t = 3.75 (upper solid red line). We include data processed by Tang *et al*. (gray symbols, right scale), which shows their finite-size extrapolated gaps for U/t = 3.75 based on the interpolation scheme proposed in figure S2 of (*1*).

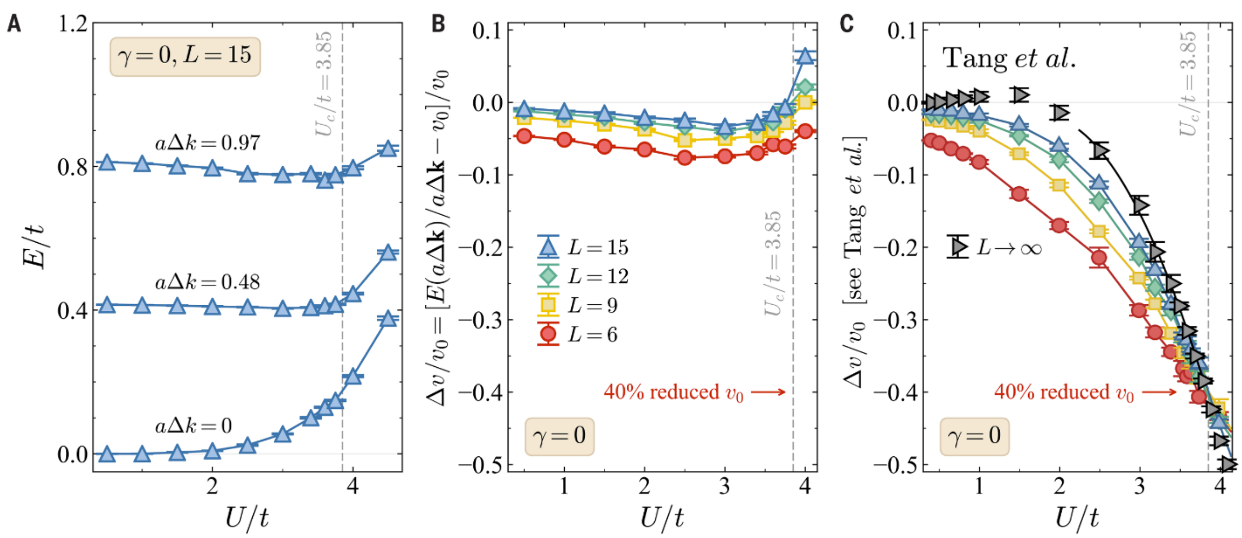


Fig. 2. Interaction effects on the low-energy excitations for the Hubbard model on the honeycomb lattice. (**A**) Dependence of the bare lowest particle-excitation energy E on the strength of the Hubbard interaction U at the Dirac point ($a\Delta k$ = 0) and at two different distances $a\Delta k$ = 0.48 and 0.97 to the Dirac point for the largest accessed linear system size L = 15 of (*1*). (**B**) Relative difference between v_0 and the rescaled lowest particle-excitation energy $E(a\Delta k)$ at the closest momentum to the Dirac point on each finite lattice, as a function of the strength of the Hubbard interaction U for different system sizes L. The red arrow indicates the 40% decrease with respect to v_0 reported in (*1*). In both panels, the dashed vertical line gives the position of the Gross-Neveu quantum critical point from (*2*). (**C**) The estimate for the renormalization of the Fermi velocity as provided by Tang *et al*., which includes the strongly finite size–affected Dirac point.

Cite as: H.-K. Tang *et al.*, *Science*
10.1126/science.aav8877 (2019).

Response to Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions"

Ho-Kin Tang,[1,2] J. N. Leaw,[1,2] J. N. B. Rodrigues,[1,2]* I. F. Herbut,[3] P. Sengupta,[1,4] F. F. Assaad,[5] S. Adam[1,2,6]†

[1]Centre for Advanced 2D Materials, National University of Singapore, 117546 Singapore. [2]Department of Physics, Faculty of Science, National University of Singapore, 117542 Singapore. [3]Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [4]School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore. [5]Institut für Theoretische Physik und Astrophysik and Würzburg-Dresden Cluster of Excellence ct.qmat, Universität Würzburg, D-97074 Würzburg, Germany. [6]Yale-NUS College, 138527 Singapore.

*Present address: Department of Physics and Institute for Condensed Matter Theory, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

†Corresponding author. Email: shaffique.adam@yale-nus.edu.sg

Hesselmann *et al.* question one of our conclusions: the suppression of Fermi velocity at the Gross-Neveu critical point for the specific case of vanishing long-range interactions and at zero energy. The possibility they raise could occur in any finite-size extrapolation of numerical data. Although we cannot definitively rule out this possibility, we provide mathematical bounds on its likelihood.

Our procedure to extract the interaction correction to Fermi velocity $\Delta v_F(k)$ from quantum Monte Carlo (QMC) simulations for on-site Coulomb interactions ($\gamma = 0$) and very close to criticality ($U \rightarrow U_c$) has been questioned (*1*). To put this in context, our figure 2 in (*2*) comprises about 120 data points for on-site interactions ranging from weak to strong coupling, and with varying long-range interaction. The critique concerns at most three of these data points. Additionally, our data can be thought of as momentum slices of figure 2, for which we use 16 such slices. Their criticism is only close to the Dirac point ($\Delta k \rightarrow 0$), and as such, it concerns only these three data points out of the ~2000 projected datasets, and therefore no more than 0.2% of the QMC data in (*2*). Most of our core findings and conclusions are unaffected. Here, we defend our claim of a Fermi velocity suppression for $\gamma = 0$, $U \rightarrow U_c$, and $\Delta k \rightarrow 0$. We consistently use the following estimator for Δv_F:

$$\Delta v_F^{\mathrm{Tang}}(k) = \lim_{L\to\infty} \frac{\Delta E(k,L) - \Delta E(0,L)}{k} \tag{1}$$

where $\Delta E(k, L)$ is obtained from QMC data for system size L. We evaluate our estimator at the smallest momentum accessible to QMC, $k_{\min} = 4\pi/\left(\sqrt{3}L_{\max}\right)$. $\Delta E(k_{\min}, L)$ is obtained for $L < L_{\max}$ from $\Delta E(k_L, L)$ and $\Delta E(0, L)$ by linear interpolation (see Fig. 1). Here $k_L = 4\pi/\left(\sqrt{3}L\right)$. Note that if we could simulate infinite lattices, our estimator would be identical to the mathematical definition of $\Delta v_F(k \rightarrow 0)$,

$$\Delta v_F^{\mathrm{Tang}}(k \to 0) = \lim_{k_{\min}\to 0} \lim_{L\to\infty} \frac{\Delta E(k_{\min},L) - \Delta E(0,L)}{k_{\min}} \tag{2}$$

Hesselmann *et al.* observe from our numerical data that close to criticality, the Dirac point energy is more strongly affected by finite lattice size than neighboring momenta. To correct for this, they outline an alternate procedure: (i) Set the Dirac point energy to zero (throwing away all information that the QMC provides about the Dirac point), and (ii) use an estimator obtained from a single system size,

$$\Delta v_F^{\mathrm{H}} = \frac{\Delta E\left(\frac{4\pi}{\sqrt{3}L_{\max}}, L_{\max}\right)}{4\pi/\left(\sqrt{3}L_{\max}\right)} \tag{3}$$

This estimator ignores any finite-size scaling information. Figure 1 shows that the two estimators give different results when applied to our QMC data. This can be understood by first noting that in the thermodynamic limit, the Hesselmann *et al.* estimator is different from the mathematical definition of the Fermi velocity,

$$\Delta v_F^{\mathrm{H}} = \lim_{k_{\min}\to 0} \frac{\Delta E\left(k_{\min}, \frac{4\pi}{\sqrt{3}k_{\min}}\right)}{k_{\min}} \neq \Delta v_F^{\mathrm{Tang}}(k \to 0) \tag{4}$$

This is illustrated graphically in Fig. 1B. Δv_F^H is taken along the black diagonal arrow, whereas $\Delta v_F^{Tang}(k \to 0)$ is taken along the red horizontal arrow. We note that if in the thermodynamic limit the two estimators disagree, then ours is always correct. However, at issue here is not the thermodynamic limit, but the finite lattice sizes achievable using QMC. Although we cannot make a priori assumptions about the functional form of $\Delta E(k, L)$ at the critical point (because we have a strongly correlated many-body state), we can still construct hypothetical functions $\Delta E_j(k, L)$ to illustrate when and why $\Delta v_F^{Tang}(k_{min})$ and Δv_F^H disagree.

First consider $\Delta E_1(k,L) = \Delta v_F^{True}(k)k + [\alpha(k)/L]$, where $\Delta v_F^{True}(k) = \Delta v_0 + \Delta v_1 k$ and $\alpha(k) = \alpha_0 + \alpha_1 k$ ($\Delta v_0 = -0.3$, $\Delta v_1 = 0.1$, $\alpha_0 = 2$, and $\alpha_1 = 1$ are all chosen to be consistent with QMC data). For $L = 24$, this gives $\Delta v_F^H \approx 0$ and $\Delta v_F^{Tang}(k_{min}) \approx \Delta v_F^{True} = -30\%$. This simple and reasonable construction shows how it is possible for the Hesselmann *et al.* estimator to find no Fermi velocity renormalization, despite there being a strong suppression correctly captured by our estimator (see Fig. 1B).

Now consider $\Delta E_2(k,L) = [\alpha_0 \delta(k)/L] + \Delta v_F^{True} k$, where $\delta(k)$ is the Dirac delta function. This is an extreme example of Hesselmann *et al.*'s concern: Only the Dirac point has finite-size effects, but no other momenta. We emphasize that this functional form is inconsistent with our numerical data. Nonetheless, for this hypothetical worst-case scenario, $\Delta v_F^H = \Delta v_F^{True}$, and $\Delta v_F^{Tang}(k_{min}) = \Delta v_F^{True} - (\sqrt{3}\alpha_0/4\pi)$. Taking $\Delta v_F^{True} = 0$ and α_0 as above, we would underestimate v_F by at most 28%.

Hesselmann *et al.*'s core claim is that "the strong suppression of the Fermi velocity … near the Gross-Neveu QCP [quantum critical point] merely reflects the enhanced finite-size effects … at the Dirac point, but not the renormalization of the actual low-energy dispersion." Because Hesselmann *et al.* cannot exclude $\Delta E_1(k, L)$ as a possible energy function, their claim is unsubstantiated. Moreover, even in the hypothetical worst-case scenario $\Delta E_2(k, L)$, for the data in Fig. 1, it would require that $\alpha_0 > 2.79$ for their claim to be correct. As seen in Fig. 1C, our QMC data lie outside the shaded region, and therefore, for this case, their claim is false. In addition, the finite-size scaling at nonzero momenta (e.g., Fig. 1D) and the observation that all the data points for $L < 24$ lie below Δv_F^H provides convincing evidence that a functional form such as $E_1(k, L)$ is more likely than $E_2(k, L)$.

Some further remarks are in order:

1) The positive Δv_F^H is physically counterintuitive, as the fermions will scatter off paramagnons and thereby slow down. This interaction with a bosonic mode is analogous to graphene interacting with phonons for which $v_F(k)$ is suppressed close to the Dirac point and enhanced for energies larger than the Debye energy [e.g., figure 2 of (*3*)]. This framework allows us to understand how a functional form such as $\Delta E_1(k, L)$ arises physically, and why Δv_F^H incorrectly gets an enhanced Fermi velocity.

2) The renormalization group flows in (*2*) were most strongly influenced by the logarithmic divergence (at finite γ) of $\Delta v_F(k)$ at large momenta, and as such, the numerical value of Δv_F at $\gamma = 0$, $U = U_c$, and $k \to 0$ is not germane to our paper. Nor did we claim to be the first to calculate it. Figure 14 of (*4*) shows a 38% suppression of $\alpha(v_F)$, which is the prefactor of the density-density correlation function (in the Brinkmann-Rice metal-insulator transition, both v_F and α vanish at the transition). Moreover, the Fermi velocity renormalization can be obtained from the specific heat ($c_v \sim T^2/v_F^2$), for which figure 13 of (*5*) calculates a ~30% enhancement of c_v at $U = U_c$. Both of these works (and ours) support a velocity suppression.

3) Another indication that $E_1(k, L)$ is more likely than $E_2(k, L)$ is the relative stability of the two estimators. We could use Δv_F^H for our QMC data, and our main conclusions would not change except at $U \to U_c$ (away from criticality, $\Delta v_F^{Tang} = \Delta v_F^H$ in the thermodynamic limit). However, (i) Δv_F^H is inconsistent with the actual QMC data, (ii) we would need different fitting procedures for different parts of our phase diagram, and (iii) Δv_F^H is unstable with changing L. Going from $L = 15$ to $L = 24$, Δv_F^H changes from –1.17% to +2.94%, whereas Δv_F^{Tang} only changes from –31.4% to –30.5%.

REFERENCES

1. S. Hesselmann, T. C. Lang, M. Schuler, S. Wessel, A. M. Läuchli, Comment on "The role of electron-electron interactions in two-dimensional Dirac fermions". *Science* **366**, eaav8877 (2019).
2. H.-K. Tang, J. N. Leaw, J. N. B. Rodrigues, I. F. Herbut, P. Sengupta, F. F. Assaad, S. Adam, The role of electron-electron interactions in two-dimensional Dirac fermions. *Science* **361**, 570–574 (2018). doi:10.1126/science.aao2934 Medline
3. A. Bostwick, T. Ohta, T. Seyller, K. Horn, E. Rotenberg, Quasiparticle dynamics in graphene. *Nat. Phys.* **3**, 36–40 (2006). doi:10.1038/nphys477
4. Y. Otsuka, S. Yunoki, S. Sorella, Universal Quantum Criticality in the Metal-Insulator Transition of Two- Dimensional Interacting Dirac Electrons. *Phys. Rev. X* **6**, 011029 (2016). doi:10.1103/PhysRevX.6.011029
5. T. Paiva, R. T. Scalettar, W. Zheng, R. R. P. Singh, J. Oitmaa, Ground-state and finite-temperature signatures of quantum phase transitions in the half-filled Hubbard model on a honeycomb lattice. *Phys. Rev. B* **72**, 085123 (2005). doi:10.1103/PhysRevB.72.085123

6. H.-K. Tang, Source data of Green's function and figure files; https://figshare.com/articles/Source_data_of_Green_s_function_and_figure_files_/5131840.

ACKNOWLEDGMENTS

Funding: Supported by the Singapore National Research Foundation (NRF-NRFF2012-01), Deutsche Forschungsgemeinschaft (SFB 1170 ToCoTronics, project C01), and NSERC of Canada. **Author contributions:** J.N.L., F.F.A., and S.A. wrote the first draft. J.N.L. designed Fig. 1 with input from all the authors. All authors discussed the results and contributed to the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data from Fig. 1 were generated in the context of (*2*) and are deposited at Figshare (*6*). In (*2*), we used a projective version of the auxiliary field quantum Monte Carlo approach that can be reproduced using the Algorithms for Lattice Fermions (ALF) open-source general implementation of the finite-temperature auxiliary field quantum Monte Carlo available at https://alf.physik.uni-wuerzburg.de.

11 November 2018; accepted 4 November 2019
Published online 6 December 2019
10.1126/science.aav8877

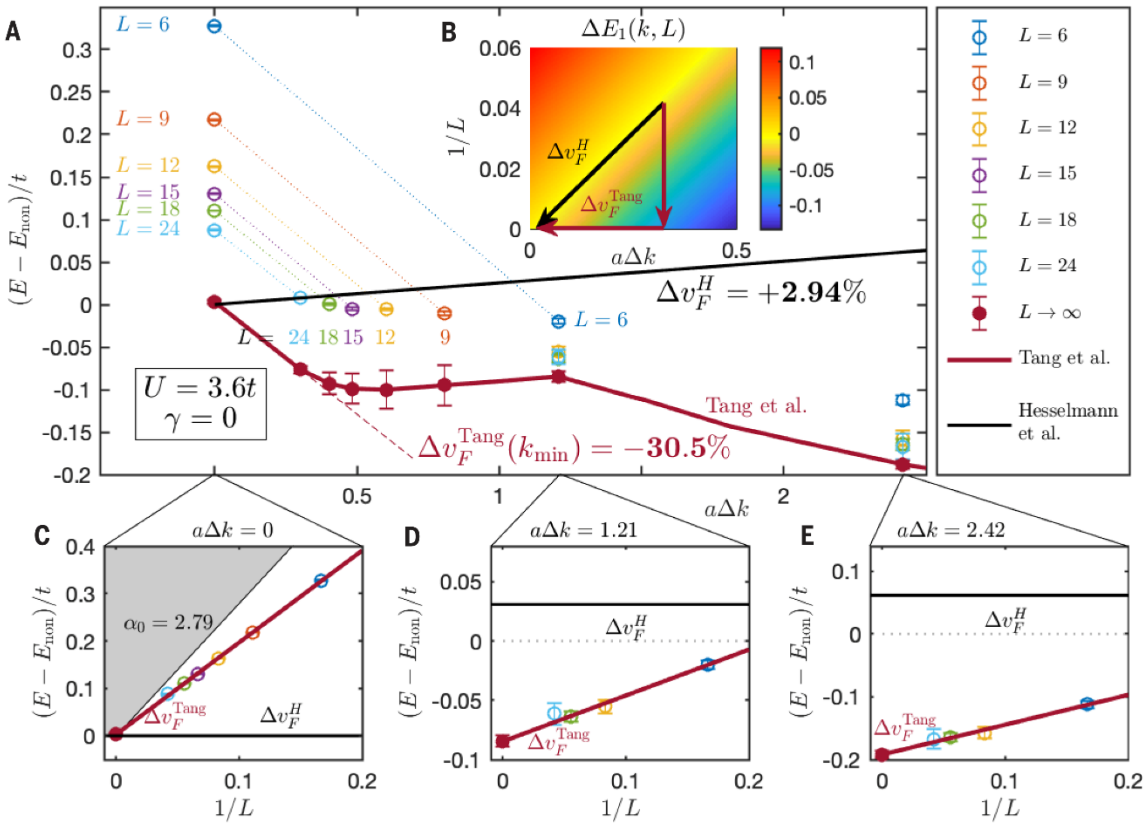
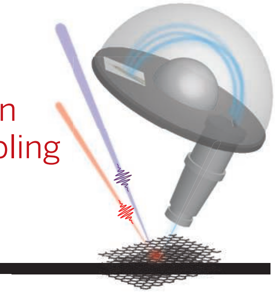


Fig. 1. Change in energy for Dirac fermions stemming from electron-electron interactions. The data were obtained using the projective quantum Monte Carlo method developed in (*2*). (**A**) Open data points are for lattice sizes $L \times L$, where L = 6, 9, 12, 15, 18, 24. The solid red line is our estimator $\Delta v_F^{\text{Tang}}(k)$ (Eq. 1), and the black line is the alternate estimator Δv_F^{H} (Eq. 3). The two estimators disagree at k_{min}: $\Delta v_F^{\text{Tang}}(k_{\text{min}})$ = −30.5% while Δv_F^{H} = +2.94%. (**B**) Our estimator is always correct in the thermodynamic limit (see text). (**C**) Finite-size scaling of the Dirac point. Because the QMC data are outside the shaded region, even in the hypothetical worst case for our estimator, our numerical data are inconsistent with an unrenormalized Fermi velocity. (**D** and **E**) Finite-size scaling of $v_F(k)$ at nonzero momenta provides clear evidence that $E_1(k, L)$ (the best case for our estimator) is more likely than $E_2(k, L)$ (the worst case for our estimator).

RESEARCH

Spectroscopic dissection of electron-phonon coupling

Na et al., *p. 1231*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

Immunofluorescence imaging of lymphatics associated with hair follicles in mouse skin

STEM CELLS

Stem cells reshape a lymphatic niche

Adult stem cells can both self-renew and regenerate new tissue upon demand. They reside in microenvironments (niches) that balance these decisions to avoid tissue overgrowth, cancer, and aging. Using murine skin as a model, Gur-Cohen *et al.* uncovered a lymphatic network of capillaries associated with the stem cell niche of hair follicles (see the Perspective by Harvey). Stem cells reshaped their lymphatic environment by switching their secretome to coordinate lymphatic-niche association. During tissue regeneration, a dynamic change in epithelial-lymphatic communication remodeled this association, synchronizing stem cell and niche behavior. —BAP

Science, this issue p. 1218; see also p. 1193

CONSERVATION ECOLOGY

Vulnerability to habitat fragmentation

Habitat fragmentation caused by human activities has consequences for the distribution and movement of organisms. Betts *et al*. present a global analysis of how exposure to habitat fragmentation affects the composition of ecological communities (see the Perspective by Hargreaves). In a dataset consisting of 4489 animal species, regions that historically experienced little disturbance tended to harbor a higher proportion of species vulnerable to fragmentation. Species in more frequently disturbed regions were more resilient. High-latitude areas historically experienced more disturbance and harbor more resilient species, which suggests that extinction has removed fragmentation-sensitive species. Thus, conservation efforts to limit fragmentation are particularly important in the tropics. —AMS

Science, this issue p. 1236; see also p. 1196

BIOCATALYSIS

Maximal efficiency from enzyme cascades

Enzymes are highly selective catalysts that can be useful for specific transformations in organic synthesis. Huffman *et al*. combined designer enzymes in a multistep cascade reaction (see the Perspective by O'Reilly and Ryan). The approach eliminates purification steps, recycles expensive cofactors, and couples favorable and unfavorable reactions. With the target molecule islatravir, an experimental HIV drug, they optimized five enzymes by directed evolution to be compatible with unnatural substrates and stable in the reaction conditions. Stereochemical purity was amplified at every enzymatic step, and the final synthesis was both atom economical and efficient. —MAF

Science, this issue p. 1255; see also p. 1199

MESOSCOPIC PHYSICS

Transmitting quantum states

The coherence of electrons in mesoscopic structures is thought to be unlikely to survive in a disordered environment. Duprez *et al*. show that this is not necessarily the case. They studied a metallic island as an example of a disordered environment. They made an electron interferometer and incorporated the island in one of the two paths through the interferometer. At sufficiently low temperatures and in the quantum Hall regime, they observed a clear interference pattern, indicating successful transmission of the

CREDITS (FROM TOP): NA *ET AL*.; E. FUCHS AND S. GUR-CHOEN/THE ROCKEFELLER UNIVERSITY

electrons' quantum state across the island. —JS

Science, this issue p. 1243

ULTRAFAST OPTICS

Tracking excitations

Illumination can be used to excite a sample from its ground state to a number of excited states. Typically, however, the details of the excitation dynamics are hidden from view because they decay so fast. Piatkowski *et al*. combined pump-probe transient absorption and two-pulse photoluminescence correlation spectroscopy, allowing them to assess stimulated emission and ground-state bleaching contributions to the transient absorption signal. This approach provides a window on the excitation dynamics within single nanocrystals and should also be useful for ultrafast nanocharacterization of complex samples. —ISO

Science, this issue p. 1240

HUMAN RETINA

Sensing light without forming images

In the rodent retina, intrinsically photosensitive retinal ganglion cells (ipRGCs) entrain circadian rhythms, modulate mood, and signal pupillary accommodation. Such responses are light-driven but not image-based. Working with donated human organ tissues, Mure *et al*. used electrophysiological approaches to identify ipRGCs in the human retina. The human retina has more cones than the retinas of nocturnal mice and rats. Differences in sensitivity, latency, and duration of responses identified three subtypes of human ipRGCs. —PJH

Science, this issue p. 1251

CANCER

A childhood tumor—from the beginning

Many adult cancers arise from clonal expansions of mutant cells in normal tissue. These premalignant expansions are defined by somatic mutations shared by the cancers. Whether pediatric cancers originate in a similar way is unknown. Coorens *et al*. studied Wilms tumor, a childhood kidney cancer. Phylogenetic analyses revealed large clones of mutant cells in histologically and functionally normal kidney tissue long before tumor development. Thus, like adult tumors, Wilms tumor appears to arise from a premalignant tissue bed. —PAK

Science, this issue p. 1247

CANCER

Finding tumor cells and killing them, too

Treatments targeting the androgen receptor are a mainstay of prostate cancer therapy. However, these treatments do not usually cure the disease and eventually lose their effectiveness. A major cause of this therapeutic resistance is the presence of neuroendocrine tumor cells, which are not sensitive to androgen inhibition. Li *et al*. found that neuroendocrine prostate cancer cells express a chemokine receptor called CXCR2. The receptor could be used to help identify these cells in tumors and represents a viable therapeutic target. —YN

Sci. Transl. Med. **11**, eaax0428 (2019).

ATMOSPHERIC SCIENCE

Here comes the flood

Atmospheric rivers (ARs) are extratropical storms that produce extreme precipitation on the west coasts of the world's major landmasses. Flood damage causes huge financial losses in U.S. West Coast communities. Now, researchers have found a possible link between ARs and flood damage. Corringham *et al.* found that increases in AR intensity and duration correspond with an estimated 10-fold increase in flood damages. The authors categorized ARs on a scale of one to five, with stages four and five representing the most intense ARs. This approach could potentially increase the efficiency of emergency preparedness for extreme flooding. —TM

Sci. Adv. 10.1126/sciadv.aax4631 (2019).

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

Worker phenotypes of genetically identical Florida carpenter ants are determined by epigenetic modification.

DENDRITIC CELLS

More DC subtypes revealed

Dendritic cells (DCs) are specialized immune cells that induce antigen-specific immunity and are sentinels for the initiation of T lymphocyte anticancer responses. Mouse DCs have traditionally been classified into two groups, cDC1 and cDC2, but Brown *et al.* discovered that the cDC2 group is made up of two previously unknown subsets called cDC2A and cDC2B. Using single-cell RNA sequencing, they found that cDC2A and cDC2B can be distinguished by expression of the transcription factors T-bet and RORγT. These subsets seem to have different functions, with cDC2A associated with wound healing and cDC2B linked to a proinflammatory state. Similar DC counterparts were found in human samples, which may provide an explanation for the mixed responses that patients have to cancer immunotherapy. —PNK

Cell **179**, 846 (2019).

NEURODEVELOPMENT

Stabilizing cell-type ratios

Brain development depends both on having enough neurons and on those neurons being connected in the right ways. Willett *et al.* found that in mice, disruption of excitatory neurons of the cerebellar nuclei, either by inactivating genes encoding specific transcription factors or by dosing the neurons with diptheria toxin, resulted in an undersized cerebellum. The size disparity was a result of fewer downstream connected cells, including granule cells, Purkinje cells, and interneurons. Despite the cerebellum having too few cells, the ratio of cell types stayed normal. Thus, the number of excitatory cerebellar nuclei neurons defines survival of a matched set of Purkinje cells. The correct ratios of cells needed to establish functional circuits are maintained by a mix of strategies that regulate both proliferation and survival of neurons. —PJH

eLife **8**, e50617 (2019).

PHOTO: GRAHAM, BERGER LAB, UNIVERSITY OF PENNSYLVANIA

MOLECULAR BIOLOGY

Epigenetic plasticity

The Florida carpenter ant, *Camponotus floridanus*, is a social insect with two worker castes that, despite sharing the same genome, are phenotypically distinct: "Major" workers defend the nest, and "minor" workers forage. Because major workers can be experimentally reprogrammed into minor workers, they are a great system to study the underlying epigenetic mechanisms for phenotypic differentiation. Glastad *et al.* compared gene expression in the brains of reprogramming-capable younger and reprogramming-incapable older major workers, as well as in the brains of major and minor workers during reprogramming. Chromatin repressor CoREST is up-regulated and required during reprogramming to repress major-biased genes, including enzymes that degrade juvenile hormone, which is key for regulating caste specificity. Thus, a chromatin-based mechanism that links transient epigenetic plasticity to long-lasting and complex social behavior has been uncovered. —SYM

Mol. Cell 10.1016/j.molcel.2019.10.012 (2019).

IMMUNOLOGY

Delicate dancing in the liver

Kupffer cells are a subset of macrophages lining the sinusoids of the liver. There, they engage in specialized functions, including the consumption and breakdown of erythrocytes. However, the interactions that imprint the tissue-specific phenotype of Kupffer cells is poorly understood. Two groups provide insights into how macrophage progenitors are directed toward this lineage. Bonnardel *et al.* and Sakai *et al.* depleted Kupffer cells and then monitored the macrophages that repopulated the liver for clues. Their findings indicate that there is a complex interplay of Kupffer cells with stellate cells, hepatocytes, and sinusoidal endothelium in the perisinusoidal space. These other cell types work to recruit monocytes and then induce and maintain the Kupffer cell phenotype via Notch–bone morphogenetic protein signaling and transforming growth factor–β family ligands. —STS

Immunity **51**, 638, 655 (2019).

MATERIALS SCIENCE

Better fatigue resistance at low cost

Framework or microarchitectured materials can be designed to have a combination of strength and stiffness, not through changes in the underlying material but through variations in lattice density and architecture. By looking at the properties of cancellous bone, which is better than compact bone for stress dampening, Torres *et al.* explored ways to enhance the fatigue properties of microarchitectured materials. They found that a key element is the proportion of materials oriented transverse to the applied loads, because these materials act as sacrificial elements during cycling loading. Thus, although there is a tendency to design framework materials to maximize the struts oriented in the direction of expected loading, this can lead to a drastic loss of fatigue resistance and only minimal decrease in the overall density. —MSL

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1905814116 (2019).

PHYSICS

Imaging heavy Dirac fermions

Most topologically nontrivial materials discovered to date have negligible electron correlations. Strongly correlated topological materials are predicted to have a rich phenomenology; however, identifying such materials and proving their topological character has been tricky. This is particularly true of samarium hexaboride (SmB_6), a compound that is theoretically expected to be a topological Kondo insulator, but the multitude of probes used to study it have yielded conflicting results. Pirie *et al.* applied scanning tunneling spectroscopy to the problem and found evidence of Dirac states with large effective masses, which is in agreement with theory. They were able to reproduce the results for multiple samples synthesized by two different growers. —JS

Nat. Phys. 10.1038/s41567-019-0700-8 (2019).

WORKFORCE

Graduate students under pressure

Although graduate student mental health is a topical issue, there is little empirical research on the subject. Hish *et al.* used a stress process model to examine the roles of mastery and social support as mediators of stress. Using a validated self-report measure of stress, data relating to stress-burnout and stress-depression relationships were collected from biomedical graduate students. Linear regression models showed that academic stressors were most predictive of burnout, whereas depressive symptoms were best predicted by family and monetary stressors. Furthermore, the relationship between stress and burnout was partially mediated by mastery and advisor support, whereas the stress-depression relationship was partially mediated by mastery, suggesting that future interventions might focus on enhancing mastery and/or improving advisor relationships. —MMc

CBE Life Sci. Ed. **18**, ar51 (2019).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

HIV VACCINES

Engineering better bnAbs

A highly effective HIV vaccine has been the goal of vaccinologists for nearly 35 years. A successful vaccine would need to induce broadly neutralizing antibodies (bnAbs) that are capable of neutralizing multiple HIV strains (see the Perspective by Agazio and Torres). Steichen *et al*. report a strategy in which the first vaccine shot can lead to immune responses that generate desired bnAbs. By combining knowledge of human antibody repertoires and structure to guide design, they validated candidate immunogens through functional preclinical testing. Saunders *et al*. designed immunogens with differences in binding strength for bnAb precursors, which enabled selection of rare mutations after immunization. The immunogens promoted bnAb precursor maturation in humanized mice and macaques. —PNK

Science, this issue p. 1216, p. 1215; see also p. 1197

ASTEROIDS

Bennu ejects material from its surface

Most asteroids appear inert, but remote observations show that a small number experience mass loss from their surfaces. Lauretta and Hergenrother *et al*. describe close-range observations of mass loss on the near-Earth asteroid Bennu (see the Perspective by Agarwal). Shortly after arriving at Bennu, navigation cameras on the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security—Regolith Explorer) spacecraft detected objects 1 to 10 centimeters in diameter moving above the surface. Analysis of the objects' trajectories showed that they originated in discrete ejection events from otherwise unremarkable locations on Bennu. Some objects remained in orbit for several days, whereas others escaped into interplanetary space. The authors suggest multiple plausible mechanisms that could underlie this activity. —KTS

Science, this issue p. 1217; see also p. 1192

VOLCANOLOGY

Caldera collapse and flank eruption

Real-time monitoring of volcanic eruptions involving caldera-forming events are rare (see the Perspective by Sigmundsson). Anderson *et al*. used several types of geophysical observations to track the caldera-forming collapse at the top of Kīlauea Volcano, Hawaiʻi, during the 2018 eruption. Gansecki *et al*. used near–real-time lava composition analysis to determine when magma shifted from highly viscous, slow-moving lava to low-viscosity, fast-moving lava. Patrick *et al*. used a range of geophysical tools to connect processes at the summit to lava rates coming out of far-away fissures. Together, the three studies improve caldera-collapse models and may help improve real-time hazard responses. —BG

Science, this issue p. 1225, p. 1212, p. 1213; see also p. 1200

SOLID-STATE PHYSICS

A timely look into electron-phonon coupling

The coupling between electrons and phonons—lattice vibrations in solids—is responsible for macroscopic quantum phenomena such as superconductivity. Yet, experimentally measuring this coupling as a function of momentum and for a particular phonon mode is tricky. Na *et al*. used time- and angle-resolved photoemission spectroscopy to excite electrons in graphite and monitor their decay, which was accompanied by the release of phonons. The time constants of these decay processes provided direct information on electron-phonon couplings in this system. —JS

Science, this issue p. 1231

QUANTUM INFORMATION

Divacancies in a diode

Solid-state defects hold great promise as the building blocks for quantum computers. Most research has focused on defects in diamond, which are difficult to integrate with existing semiconductor technologies. An alternative two-vacancy neutral defect in silicon carbide (SiC) has a long coherence time but suffers from broad optical linewidths and charge instability. Anderson *et al*. fabricated these defects in a diode made out of commercially available SiC. Reverse voltage created large electric fields within the diode, tuning the frequencies of the defect's transitions by hundreds of gigahertz. The electric fields also caused charge depletion, leading to a dramatic narrowing of the transitions. The technique should be readily generalizable to other quantum defects. —JS

Science, this issue p. 1225

STRUCTURAL BIOLOGY

Regulating synaptic signals

In the brain, AMPA-type glutamate receptors (AMPARs) are ion channels that play key roles in synaptic plasticity, cognition, learning, and memory. Two classes of subunits, the claudin family and the cornichon family, regulate AMPAR gating and trafficking. Previous structures have been presented of AMPAR bound to claudin homologs. Now, Nakagawa reports a high-resolution structure of AMPAR bound to the cornichon homolog CNIH3, determined by cryo–electron microscopy (see the Perspective by Schwenk and Fakler). In contrast to a predicted topology of three transmembrane helices and an intracellular amino terminus, CNIH3 has four transmembrane helices, and both the amino and carboxyl termini are extracellular. The structure reveals the architecture of the interaction interface between AMPAR and CNIH3 and suggests a role for lipids in regulating the assembly and function of the AMPAR-CNIH3 complex. —VV

Science, this issue p. 1259; see also p. 1194

PHYSIOLOGY

Translating into a bigger pancreas

The messenger RNA translation factor eIF5A promotes cell proliferation during development and in tumors. This ability depends on eIF5A hypusination, a posttranslational modification specific to eIF5A. Levasseur *et al.* found that eIF5A hypusination was critical for postnatal expansion of pancreatic β cell mass. Mice that could not perform hypusination in β cells did not produce sufficient cyclin D2 to sustain cell cycling and developed diabetes in response to diet-induced obesity. Thus, eIF5A hypusination enables β cells to proliferate, which is required to increase insulin production and maintain glucose homeostasis. —WW

Sci. Signal. **12**, eaax0715 (2019).

IMMUNODEFICIENCIES

Putting JNK1 on the immunodeficiency map

Impaired T helper 17 cell immunity is the shared element among a group of inherited immunodeficiencies that are associated with chronic mucocutaneous candidiasis (CMC). Li *et al*. studied three patients from a single family who had a combination of CMC and an atypical

form of connective tissue disorder that had some features of Ehlers-Danlos syndrome. Whole-exome sequencing identified a loss-of-function splice-site mutation in the *MAPK8* gene encoding c-Jun N-terminal kinase 1 (JNK1) that causes JNK1 haploinsufficiency with autosomal dominant inheritance. The complex clinical phenotype in these patients results from defects in signaling downstream of both interleukin-17 and transforming growth factor–β (TGF-β) cytokines. Thus, JNK1-mediated signaling plays a critical role in maintaining normal immunity to *Candida*, and TGF-β promotes homeostasis of connective tissues. —IW

Sci. Immunol. **4**, eaax7965 (2019).

RESEARCH ARTICLE SUMMARY

VOLCANOLOGY

The tangled tale of Kīlauea's 2018 eruption as told by geochemical monitoring

Cheryl Gansecki*, R. Lopaka Lee, Thomas Shea, Steven P. Lundblad, Ken Hon, Carolyn Parcheta

INTRODUCTION: Fissures sliced through Kīlauea Volcano's lower east rift zone on 3 May 2018, eventually engulfing hundreds of structures in lava flows and triggering a collapse at the summit. During the eruption, we employed a rapid routine for geochemical analysis of lava, developed over 6 years of monitoring the prior continuous eruption at Kīlauea. The application of this routine elevated lava chemistry to a near real-time data stream in eruption monitoring, similar to seismic and geodetic data. It provided an unparalleled opportunity to understand changes in magma characteristics during a rapidly evolving eruptive crisis.

RATIONALE: Lava chemistry provides vital information on the underground sources of magma, eruptive conditions, temperature, and physical properties of lava flows. However, analytical techniques are typically slow, leaving chemical analysis of lava as a retrospective tool in the volcano sciences. We developed an analytical procedure to characterize the geochemistry of lava within a few hours of sample collection, allowing us to identify a specific suite of major and trace elements that track lava compositions and estimate lava temperatures through chemical geothermometers. This information was used to inform response teams of shifts in eruptive conditions.

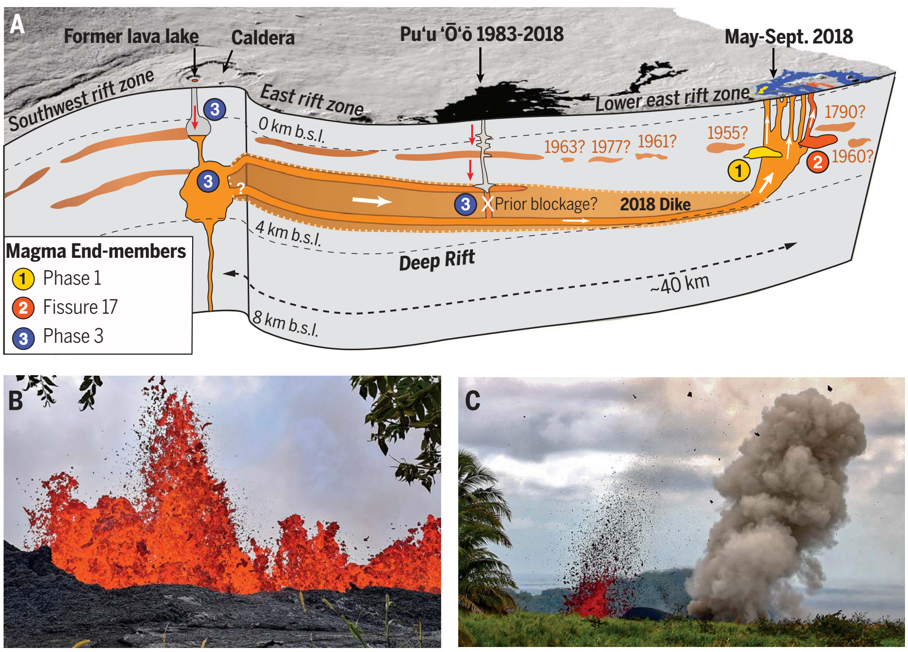
RESULTS: The initial fissures erupted low volumes of chemically evolved basaltic lavas from 3 to 9 May, which were viscous and cool (~1110°C). On 13 May we detected less-evolved compositions and an increase in inferred lava temperatures (~1130°C). We informed science and response teams that the arrival of more fluid and voluminous lava was likely. Beginning 17 to 18 May, the lava from the primary fissures became increasingly less chemically evolved, hotter, and more fluid. By 28 May, activity focused on a single vent (fissure 8). This vent fed a massive outpouring of hotter (~1145°C) lava that continued for more than 2 months. During this stage, lavas became slightly hotter and lost the cargo of lower-temperature minerals that were initially abundant. The lava carried olivine crystals with unusually high MgO, indicative of the presence of much hotter magma (>1270°C) somewhere in the plumbing system. A second dominant olivine population formed in cooler magma similar to what was being erupted previously at the summit lava lake.

We also identified simultaneous, but more explosive, repetitive outbursts of andesite lava. This highly viscous and evolved composition, not previously known from Kīlauea, erupted at low temperatures (1060° to 1090°C) on a fissure offset from the other eruption fissures. The chemical and mineralogical fingerprint of this lava was also detected at other fissures several kilometers from the andesite vent.

CONCLUSION: Analysis of the data during the eruption revealed that at least three different sources of magma were feeding the eruption. The first two were the chemically evolved basalt of the initial fissures and the highly viscous andesite. Both are volumetrically minor sources that represent distinct pockets of old residual magma from Kīlauea's east rift zone that evolved for more than 55 years, cooling and crystallizing at depth. The third and volumetrically more substantial source was less-evolved and hotter basalt of fissure 8. This source was similar in composition to the magma erupted at Kīlauea in the years before 2018 and was ultimately derived from the summit region. Draining and collapse of the summit by this voluminous eruption may have stirred up deeper, hotter parts of the summit magma system and sent mixed magma down the rift. By the final 20 days of the eruption, most magma stored within the active rift system had flushed out. Post-eruption analyses done by traditional geochemical methods confirmed that the rapid-response routine produced comparable data and validated the models proposed during the active eruption. Our work has demonstrated that geochemical analyses of lava samples in near-real-time can yield critical information that enhances hazard assessments and risk mitigation during an eruption. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaz0147



The 2018 lower east rift zone eruption of Kīlauea Volcano with inferred magma sources and pathways. (**A**) Simplified model of Kīlauea's magma system feeding the 2018 lower east rift zone eruption and locations of hypothesized magma end-members (b.s.l., below sea level). (**B**) Fluid basalt erupting from fissure 20 on 20 May 2018. (**C**) Fissure 17 erupting andesite more explosively 800 m away. Photos by U.S. Geological Survey.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: gansecki@hawaii.edu
Cite this article as C. Gansecki *et al.*, *Science* 366, eaaz0147 (2019). DOI: 10.1126/science.aaz0147

RESEARCH ARTICLE

VOLCANOLOGY

The tangled tale of Kīlauea's 2018 eruption as told by geochemical monitoring

Cheryl Gansecki[1]*, R. Lopaka Lee[2], Thomas Shea[3], Steven P. Lundblad[1], Ken Hon[1], Carolyn Parcheta[2]

Changes in magma chemistry that affect eruptive behavior occur during many volcanic eruptions, but typical analytical techniques are too slow to contribute to hazard monitoring. We used rapid energy-dispersive x-ray fluorescence analysis to measure diagnostic elements in lava samples within a few hours of collection during the 2018 Kīlauea eruption. The geochemical data provided important information for field crews and civil authorities in advance of changing hazards during the eruption. The appearance of hotter magma was recognized several days before the onset of voluminous eruptions of fast-moving flows that destroyed hundreds of homes. We identified, in near real-time, interactions between older, colder, stored magma—including the unexpected eruption of andesite—and hotter magma delivered during dike emplacement.

Chemical analysis of lava provides a wealth of information about physical properties of flows, eruptive conditions, magma transport, and magma storage. Typically, analyses trail events by weeks to months, which inhibits combining geochemical data with live streams of seismic, geodetic, gas chemistry data, and field observations. Chemical changes have particular significance at Kīlauea Volcano, where the initial phases of many fissure eruptions are dominated by differentiated lava from past eruptions that is stored in the rift zone [e.g., (*1*–*3*)]. The degree of fractionation, the volume of stored magma, and the amount of mixing with intruding magma are highly variable and exert substantial control on eruption behavior. In long-lived rift eruptions such as Puʻu ʻŌʻō (1983–2018), fractionated magma may take several years to flush out, before lava compositions become dominated by the newer magma and stabilize (*4*). During the 2018 Kīlauea lower East Rift Zone (LERZ) eruption, the collection and rapid chemical analysis of lava samples were key to identifying and monitoring properties of lava and deciphering eruptive processes as they occurred.

The U.S. Geological Survey (USGS) Hawaiian Volcano Observatory (HVO) and the University of Hawaii at Hilo (UH Hilo) have partnered since 2012 to develop a rapid analytical protocol to characterize lava from active eruptions within a few hours of collection. The procedure developed during the continuous eruption of Puʻu ʻŌʻō rapidly analyzes a limited suite of trace elements and major elements to identify changes in maximum lava temperatures, crystal fractionation trends, storage, and source origins. The chemical data are easily integrated with other monitoring data sources during an eruptive crisis, allowing interpretations and hazard assessments in real time not previously possible.

2018 Kīlauea eruption

The collapse of the long-lived Puʻu ʻŌʻō vent of Kīlauea volcano on the Island of Hawaiʻi on 30 April was followed by downrift propagation of a dike into the LERZ (*5*). On 3 May, eruptive fissures opened in the Leilani Estates subdivision west of Pohoiki Road. Fifteen fissure segments erupted along a single fracture trend during the first week, followed by an eruptive pause beginning on 9 May, although local earthquakes and deformation continued (*5*). On 12 May, new fissures opened east of Pohoiki Road downrift along the same trend, eventually extending 6.8 km in total length along 24 fissures (Fig. 1) (*5*). Only one fissure, 17, opened offset from the others (Fig. 1). Lava output increased rapidly as the main eruptive activity shifted back uprift through May. Massive flows from fissure 8 commenced on 28 May and remained vigorously active until 4 August (*5*). The last active lava was observed in the fissure 8 vent on 5 September. Effusion rates up to 200 m^3/s (dense-rock equivalent) produced an estimated total of 0.8 to 1.0 km^3 of lava, making this the largest LERZ eruption in 200 years (*5*).

The timing and pattern of 2018 LERZ fissure eruptions define four general eruptive phases. The first three eruptive phases coincide largely with the changes in composition that we describe below, whereas phase 4 produced no lava to analyze. We also consider the eruption of fissure 17, and the later reactivation of certain fissures (particularly 13, 18, 19, and 22), as discrete events that warrant separate consideration given their compositional differences.

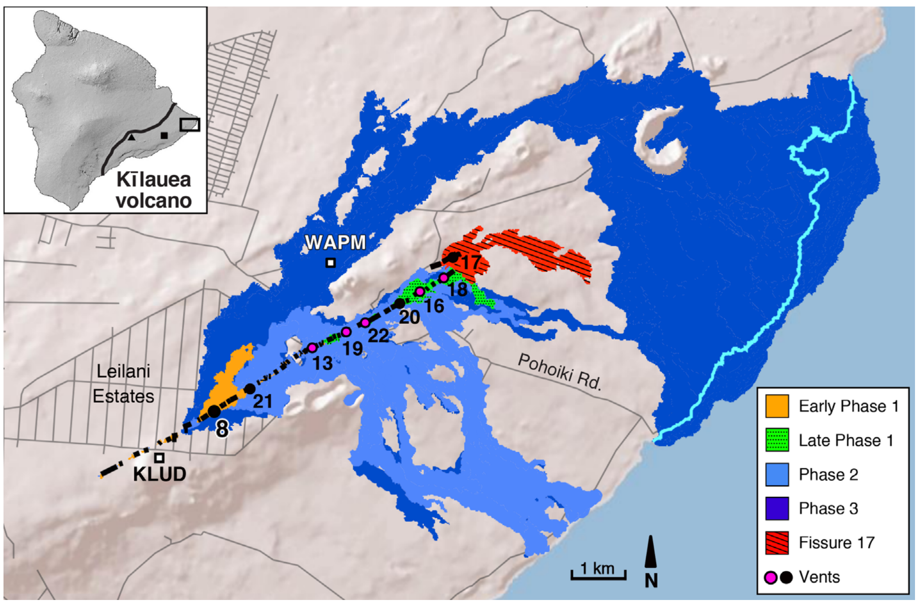


Fig. 1. Map of lava flows and vents from the 2018 Kīlauea eruption. Flows are color-coded by eruptive sequence: gold is lava from early phase 1 (initial fissures), stippled green is late phase 1 (early downrift fissures), striped red is fissure 17, light blue is phase 2 (flows of 17 to 27 May), dark blue is phase 3 (after 27 May). Numbers mark vent locations mentioned in text; magenta dots are reactivated fissures. The pre-eruption coastline is shown in light blue. Locations of tilt station WAPM and seismic station KLUD are shown as squares. Inset map shows the Island of Hawaiʻi and Kīlauea Volcano with the summit Halemaʻumaʻu lava lake (black triangle) and Puʻu ʻŌʻō vent (black square) marked; map area is outlined. Map data: Esri, County of Hawaii, Hawaii Statewide GIS Program and USGS (*25*).

[1]Department of Geology, University of Hawaiʻi at Hilo, 200 W. Kawili Street, Hilo, HI 96720, USA. [2]U.S. Geological Survey, Hawaiian Volcano Observatory, 1266 Kamehameha Avenue, Suite A8, Hilo, HI 96721, USA. [3]Department of Geology and Geophysics, SOEST, University of Hawaiʻi at Mānoa, Honolulu, HI 96822, USA.
***Corresponding author. Email: gansecki@hawaii.edu**

Early phase 1 (3 to 9 May, fissures 1 to 15, ~0.1% of total erupted volume) and late phase 1 (12 to 18 May, fissures 16, 18 to 20, 22, ~0.1% of total erupted volume) were both small-volume, short-lived fissures separated by a short eruptive pause and a shift downrift (Fig. 1). Phase 2 saw increased lava effusion that sent large flows to the southeast as fountaining migrated back uprift (17 to 27 May; fissures 21 to 24, others continue or reactivate, ~3 to 7% of total volume). During phase 3, voluminous output from fissure 8 vent produced a large and destructive channelized lava flow (28 May to 4 August; ~92 to 96% of total volume). Fissure 17 produced fountains, highly viscous lava flows, and explosive bursts (13 to 25 May; ~0.5 to 0.6% total volume).

Geochemical analysis

USGS-HVO field crews collected 113 molten or recently solidified samples during the eruption. Samples were delivered 2 to 12 hours after collection to UH Hilo, where they were dried, given a quick petrographic overview, and prepared for analysis (*6*). We analyzed all samples by energy-dispersive x-ray fluorescence spectroscopy (EDXRF) for a limited suite of whole-rock major (Ca, K, Ti, Mg) and trace (Rb, Sr, Zr, Y, Nb) elements (data S1) (*7*). The main advantages of EDXRF for geochemical monitoring are minimal sample preparation and rapid data production. The turnaround time from rock to data was 1 to 2 hours, and we analyzed most within 24 hours of field collection. We analyzed a subset of samples during and after the eruption by conventional wavelength-dispersive X-ray fluorescence (WDXRF) spectroscopy for a full suite of elements (data S2) (*7*). We also determined matrix glass (data S3) (*7*) and phenocryst (data S4) compositions by electron microprobe analysis (EMPA) on polished thin sections and grain mounts from representative samples. Matrix glass compositions represent the melt component of magma.

Geothermometry

We calculated lava temperatures using Ca- or Mg-based geothermometers calibrated using Kīlauea glasses (*7*, *8*). For initial estimates of lava temperature, we used the geothermometers on whole-rock, EDXRF-derived CaO or MgO. Because the equations were empirically derived from mineral-free glass, calculated temperatures will be biased high when Ca- or Mg-rich minerals are present and thus represent a maximum temperature (given as T_{CaOmax} and T_{MgOmax}). These estimates generally agreed within ~5°C; greater divergence is an indication of disequilibrium between the glass and mineral phases (*9*). EMPA analysis of matrix glass for MgO and CaO produces more accurate melt temperatures, but takes longer to obtain. T_{CaOG} and T_{MgOG} were generally 5° to 15C° cooler than T_{CaOmax} and T_{MgOmax} (Table 1) unless the sample was crystal-poor. Some samples showed larger differences, in particular the highly evolved and crystal-rich andesite from fissure 17.

Table 1. Calculated average eruptive temperatures and representative incompatible-element compositions. Whole-rock Zr and Nb compositions, plus average whole-rock and glass temperatures from CaO and MgO glass thermometers (*8*). Bulk maximum temperatures from rapid-response EDXRF and WDXRF duplicates (in parentheses) generally agree within 2° to 3°C (data S1 and S2) (*7*). Glass temperatures from EMP analysis of MgO and CaO in matrix glass (data S3). Glass temperatures are lower than the bulk-rock temperatures, except in later phase 3 where few Ca-bearing crystals are present. The rapid-analysis temperatures show the same pattern as the glass: substantially cooler than Puʻu ʻŌʻō in early phase 1, higher temperature in late phase 1, then reaching Puʻu ʻŌʻō-like temperatures in phase 3.

Source	Bulk rock (max) temp.		Glass temp.		Zr (ppm)	Nb (ppm)
	T_{CaO} (°C)	T_{MgO} (°C)	T_{CaO} (°C)	T_{MgO} (°C)		
Halemaʻumaʻu lava lake	N/A*	(1160)	N/A	1155	140 (136)	12 (12)
Puʻu ʻŌʻō 2018	1149 (1152)	1148 (1154)	1144	1143	145 (138)	14 (13)
2018 LERZ Eruption:						
Early Phase 1	1111 (1113)	1107 (1111)	1104	1097	288 (283)	28 (27)
Late Phase 1	1128 (1131)	1130 (1132)	1113	1106	222 (213)	20 (19)
Phase 2	1144 (1143)	1158 (1160) †	1129	1126	169 (166)	16 (15)
Phase 3	1143 (1144)	1180 (1176) †	1142	1146	153 (146)	14 (13)
Fissure 17 explosive	1062 (1063)	1068 (1065)	1028	1033	543 (540)	41 (41)
Fissure 17 late flow	1093 (1098)	1088 (1096)	1079	1071	367 (365)	31 (31)
Reactivated F22	1097 (1103)	1093 (1108)	1084	1074	348 (329)	30 (28)
Reactivated F13	1115 (1117)	1114 (1119)	1094	1093	291 (284)	25 (24)

*Olivine only, T_{CaO} not applicable. †Entrained (high-MgO) olivine makes temperature estimates too high.

Results

Early phase 1: Eruption of low-temperature, highly differentiated lavas (3 to 9 May 2018)

Highly differentiated basalt erupted from 15 fissures in or near the Leilani Estates subdivision during the first week of the eruption. Cooling and crystallization produce differentiated or evolved magmas enriched in "incompatible" elements (Zr, Nb, K, Ti, and others) not incorporated into crystals. Early phase 1 incompatible–trace-element concentrations (Zr and Nb) were twice that of average lava erupted during the past 35 years of activity on Kīlauea (Table 1, Fig. 2, and data S1). Compatible major elements MgO (Fig. 2) and CaO (Fig. 3) were low, giving a maximum temperature of ~1110°C, about 30° to 40°C lower than temperatures typical of previous Puʻu ʻŌʻō and Halemaʻumaʻu eruptions (Table 1). TiO_2 concentrations were much higher in this phase than any others (Fig. 4).

The whole-rock WDXRF analyses of SiO_2 [51.7 to 52.3 weight % (wt %)] and alkalis (Na_2O+K_2O = 3.8 to 4 wt %) confirmed values higher than typical Kīlauea compositions (*10*) and were classified as silica-rich tholeiitic basalt to borderline basaltic andesite (data S2).

We calculated representative melt temperatures of 1104°C from CaO and 1097°C from MgO on matrix glasses, which were about 10°C cooler than maximum temperatures calculated from EDXRF data. Glass SiO_2 compositions from the 2018 LERZ, as well as earlier Puʻu ʻŌʻō and summit samples, were mostly typical for Kīlauea tholeiitic basalts, averaging 50 to 51 wt %. Higher glass SiO_2 was only observed in two phase 1 samples, in addition to samples from fissure 17 and several reactivated fissures (data S3).

Rapidly cooled lava from early phase 1 was glassy and appeared mostly aphyric in hand specimen, but we identified 30 to 40% microlites with scanning electron microscopy (SEM). These crystals were mostly high–aspect ratio plagioclase laths 20 to 100μm in length. Larger, more equant phenocrysts of plagioclase and pyroxene up to 1 mm in size were also present. Plagioclase of An_{52-65} was most common, though a few had more anorthitic cores (Fig. 5). Pyroxenes were mostly augite with some high-Mg orthopyroxene-pigeonite (~70% enstatite component) (fig. S1). Most phenocrysts showed strong normal zoning with more Mg-rich cores. Rare ilmenite and Fo_{64-70} olivine (Fig. 5) were present in some samples.

Late phase 1: Eruption of higher-temperature, less-differentiated lavas (12 to 18 May 2018)

Tremor and deformation continued through the 9 to 11 May pause in eruptive activity, tracking migration of the dike 2 km eastward over 3 days (Fig. 3) (*5*). Beginning on 12 May, fissures 16, 18 to 20, and 22 erupted east of Pohoiki Road, downrift of the earlier fissures,

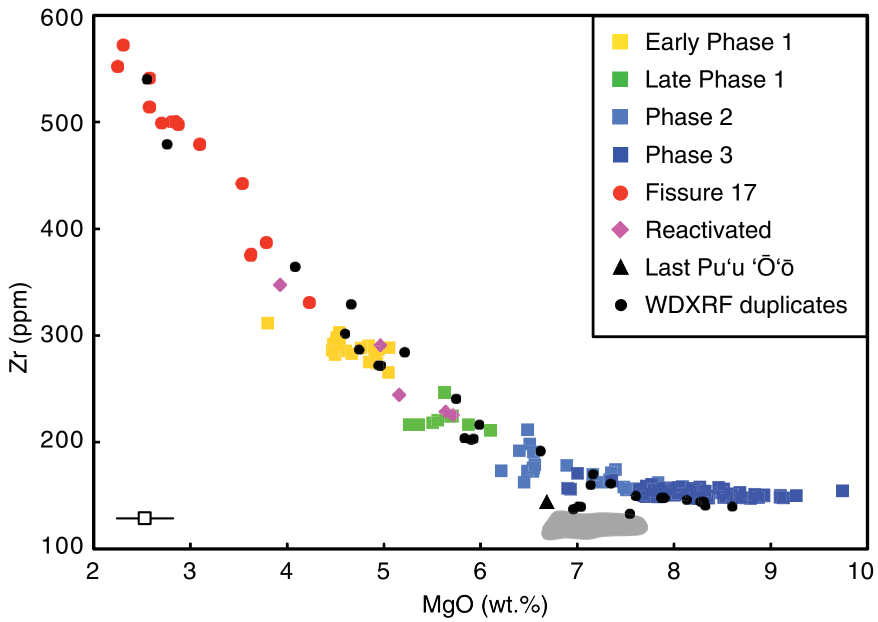


Fig. 2. Plot of incompatible element (Zr) versus a differentiation index (MgO) for all lava samples from the 2018 Kīlauea eruption. Data from whole-rock EDXRF (data S1) (7) except black circles are WDXRF data obtained on a subset of the same samples (data S2) (7). Datasets show good agreement despite relatively large error for MgO. Gray area is region of 2016–2018 values from Pu‘u ‘Ō‘ō lava; black triangle is the last lava from Pu‘u ‘Ō‘ō. Bar in lower left gives estimated EDXRF 1 SD MgO error of ±0.31 wt % (Zr 1 SD error of ±3 ppm is within symbol size).

but along the geophysically defined strike of the intrusion (Fig. 1).

The composition of the late phase 1 lava was promptly recognized as substantially less fractionated than the initial western fissure lava, having lower incompatible elements (Fig. 2 and Table 1), higher CaO (Fig. 3), and higher MgO (Fig. 2). The T_{CaOmax} of 1128°C was 15° to 20°C higher than early phase 1. MgO and CaO glass compositions yielded average temperatures of 1106° and 1113°C, respectively. Though hotter than the early phase 1 eruptions, late phase 1 lava was still 20° to 30°C cooler and less mafic than prior Pu‘u ‘Ō‘ō or Halema‘uma‘u lava (Table 1) (*11*).

Late phase 1 rocks contained 25 to 30% small (1 to 3 mm) plagioclase and pyroxene phenocrysts, with fewer 10- to 50-µm microlites in the glassy spatter compared to early phase 1 samples. Plagioclase compositions were mostly An_{55-65} from the first fissures in this sequence, increasing to An_{70} a few days later. Pyroxenes were mostly augite, and we identified only one orthopyroxene phenocryst. Intergrown clots of plagioclase and augite were common, and most of the pyroxenes showed sector and/or concentric zoning. Minor olivine of Fo_{68-75} was also present, reaching Fo_{78-79} in an early 18 May sample. Fissure 19, though active on 17 May, had anomalously high K_2O and low TiO_2, similar to some of the later reactivated fissures (Fig. 4).

Phase 2: Higher-temperature mafic mixed lava and increased effusion (17 to 27 May 2018)

We identified an abrupt change in chemistry on the afternoon of 17 May, when fissure 21 erupted more mafic lava in the middle of the western fissures in Leilani Estates. The following day, fissure 20 lava in the eastern group shifted from typical late phase 1 lava with ~6 wt % MgO and 225 parts per million (ppm) Zr in the morning, to 6.5 wt % MgO and 198 ppm Zr by evening of the same day (data S1) (*7*).

Between 18 and 26 May, effusion rates increased markedly at fissures 16, 18 to 20, and 22. Vents also became active again uprift, particularly fissures 6, 13, and 15 (Fig. 6). By 19 May, large, fast-moving lava flows from fissures 20 and 22 reached the ocean. We used EDXRF analysis to calculate T_{CaOmax} of 1138° to 1150°C (Fig. 3). Zr and MgO values approached Pu‘u ‘Ō‘ō-like compositions, but T_{MgOmax} was anomalously high (Fig. 2 and Table 1). This was likely due to entrainment of high-MgO olivine crystals, out of equilibrium with the melt (*9*). Glass CaO and MgO temperatures agreed, averaging 1128°C.

The bulk composition of phase 2 lava was substantially less differentiated than previous samples. Whole-rock MgO ranged from 7 to 8 wt %, Zr from 175 to 150 ppm, and glass MgO increased from 4.7 to 6 wt % (Fig. 5). Considering the widespread spatial distribution of phase 2 vents, compositional variability was moderate and most strongly correlated with time.

Phenocrysts, by contrast, ranged widely in composition, with plagioclase cores of up to An_{78} and rims as low as ~An_{60}. Olivine compositions ranged from Fo_{69} (rims) to Fo_{89} (cores), the latter values surpassing any from Pu‘u ‘Ō‘ō or recent summit eruptions (Fig. 5). Pyroxene phenocrysts were mostly augite, but some pigeonite was also present, mostly in crystal cores (fig. S1). Cr-spinel inclusions, more common in Halema‘uma‘u than Pu‘u ‘Ō‘ō samples [e.g., (*11*)], were found in some of the high-Fo olivine phenocrysts.

Phase 3: Voluminous eruption of high-temperature mafic lava (28 May to 4 August 2018)

On 28 May, a massive outpouring of basalt began from fissure 8. These high-volume flows bear similarities in bulk composition, phenocryst assemblages, and temperature to Pu‘u ‘Ō‘ō and summit lavas. Most element concentrations leveled out at values similar to those of recent tholeiitic basalt at Kīlauea (Figs. 2 and 3 and Table 1). T_{CaOmax} remained at 1142° to 1147°C for the rest of the eruption, with the exception of several reactivated fissures discussed below. Whole-rock MgO continued to vary, but increased to up to 9 wt %, higher than the ~7 wt % seen in the previous 10 years at Pu‘u ‘Ō‘ō (Fig. 2), likely due to entrainment of high-MgO olivine. SiO_2 (~51 wt %), alkalis (~2.7 wt %), and most other major elements were similar, if proportionately lower than values of basalts erupted from Pu‘u ‘Ō‘ō (data S2) (*7*).

We confirmed the presence of Mg-rich (Fo_{88-89}) olivine cores and crystals using the electron microprobe. Rims of most of the olivine crystals had lower Fo (Fo_{78-80}), closer to equilibrium with glass MgO concentrations averaging 6.3 wt % through the end of the eruption. These rim and glass values were similar to Pu‘u ‘Ō‘ō compositions (Fig. 5). Glass temperatures were within a few degrees of T_{CaOmax} in most of these plagioclase-poor lava flows, with the exception of a few crystal-rich distal samples from late May and June (Fig. 5 and Table 1).

The phenocryst cargo of the lava changed from augite, plagioclase, and olivine, to near olivine-only in late June. Cr-spinel was more common in the later samples, mostly as inclusions in or in close association with high-Fo olivine phenocrysts. Plagioclase and pyroxene microlites were present in quenched samples, though in smaller proportion than in the earlier phases of the eruption.

Fissure 17: Eruption of highly evolved lava (13 to 25 May 2018)

Fissure 17 erupted through late phase 1 and 2 from an en echelon segment offset north of the main fissure trend and at the distal

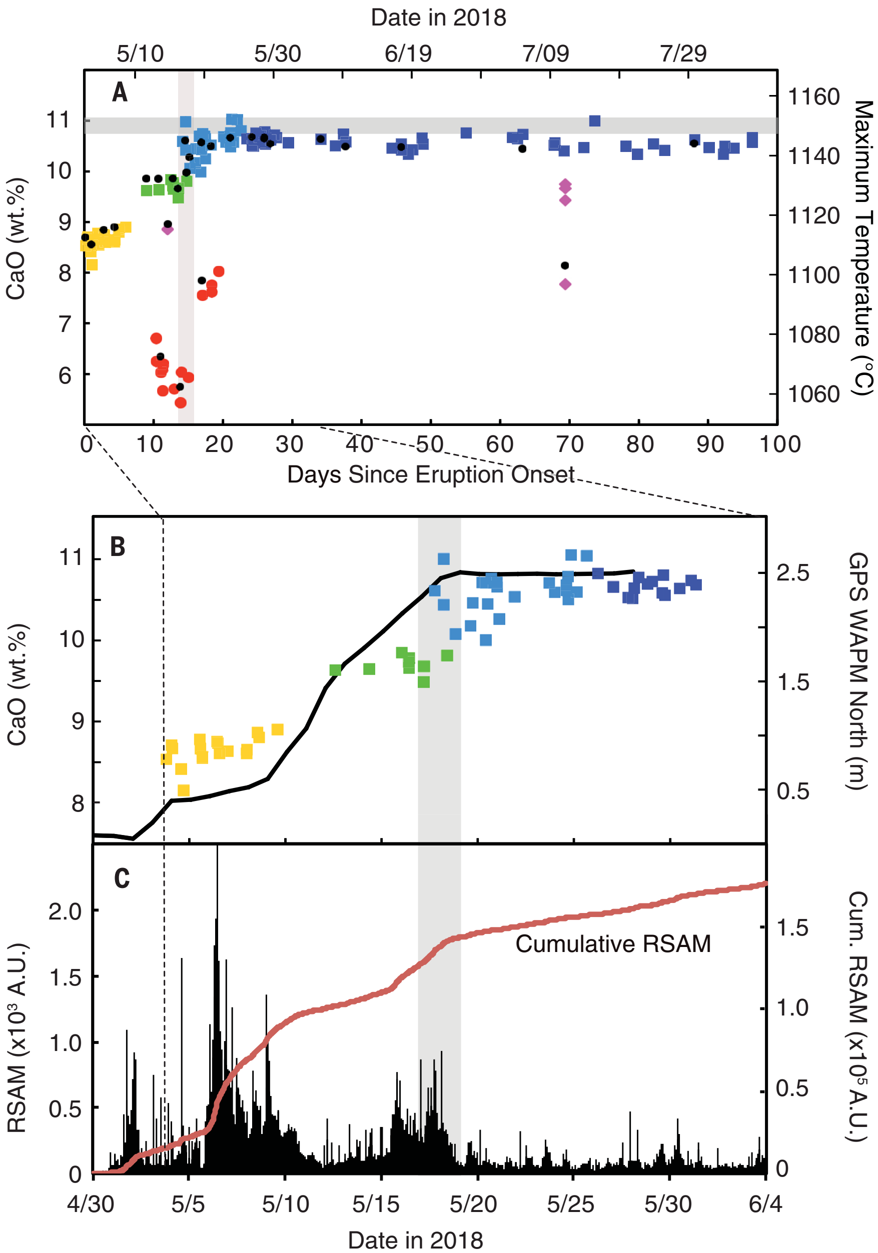


Fig. 3. Temporal variation of bulk CaO and maximum temperature and correlation with nearby deformation and seismicity. (**A**) Plot of whole-rock CaO wt % and calculated temperature variation over time during the 2018 eruption. Whole-rock CaO provides maximum temperature estimate using CaO-in-glass thermometer of (*8*). Data from EDXRF, black circles are duplicate analyses by WDXRF. Horizontal gray bar shows average Puʻu ʻŌʻō CaO composition over the previous 3 years (10.9 ± 0.2 wt %). Symbols as in Fig. 2. (**B**) Whole-rock CaO wt % as above superimposed on the northward motion of WAPM GPS station (right axis, black line). Dashed line is eruption onset, vertical gray bar indicates inferred arrival time of main intrusion. (**C**) Vertical black bars show 1-hour average of RSAM [Real-Time Seismic Amplitude Measurement (*26*) in arbitrary units A.U. on left axis] at the KLUD station. Red line is cumulative hourly RSAM on right axis. (B) and (C) are plotted over the period 30 April to 4 June 2018. Station locations shown in Fig. 1; WAPM was destroyed on 29 May.

northeastern end (Figs. 1 and 6). This was the only vent substantially offset (~200 m, Fig. 1) from the rest of the fissure system. The composition of this lava is more evolved than any previous known eruptions on Kīlauea Volcano, either from historic or prehistoric flows [e.g., (*10*)]. Zr concentrations were >500 ppm, and MgO was <3 wt % (Fig. 2 and Table 1). T_{CaOmax} reached low values of 1062° to 1093°C.

Whole-rock SiO_2 was 55 to 60 wt % with total alkalis of 4.5 to 5.7 wt % (data S2) (*7*), classifying it as an andesite. The lava apparently became more evolved during the first few days of activity, then quickly changed to a less evolved, basaltic andesitic composition (Fig. 3). Matrix glass from fissure 17 andesite attained 67 wt % SiO_2 (data S3) near the end of late phase 1. Initial CaO and MgO glass temperatures were ~1050°C, dropping to 1030° to 1035°C in high-silica spatter samples, then increasing to 1070° to 1080°C in the last-erupted lava.

Phenocrysts were predominantly plagioclase and two pyroxenes, all with wide compositional ranges (Fig. 5 and fig. S1). Plagioclase and pyroxene microlites were also abundant. Unlike other lavas from the eruption, at least a few fissure 17 samples contained ilmenite, titanomagnetite, apatite, or rare inclusions of pyrrhotite, but no olivine. Inclusions, crystal clusters, and large, coarse-grained, glassy enclaves were common.

The western end of fissure 17 erupted more explosively than other vents. It is unclear how much of the explosivity was due to the high viscosity of the magma and how much was due to interaction with external groundwater, as the biggest explosions occurred on the west end, where the volumetric output was low and viscosity highest. Similar explosive behavior was not seen at other low-volume but less-differentiated fissures nearby. Activity at the central fissure 17 vent was predominantly fissure fountaining producing lava flows, simultaneous with explosions at the west end. The extremely viscous flow, with its high density and yield strength and intense radiative heat, made sampling challenging. Most spatter samples were collected 200 to 300 m from the explosive vent.

Reactivation of fissures: Eruption of mixed compositions

Many of the eruptive vents along the fissure system reactivated intermittently days to weeks after their initial cessation (Figs. 1 and 6). Several showed a change to more evolved rather than less evolved compositions, at least briefly. Fissure 13 reactivated on 15 May after being inactive for 6 days and produced lava with a bulk composition similar to that of its initial eruption, but with much higher matrix glass SiO_2 (54 to 55 wt % versus 49 to 51 wt %) and apatite, which was not found in any of the initial eruptions. Fe-rich orthopyroxene phenocryst compositions from this sample overlapped with fissure 17 compositions (fig. S1) and were unlike any other unit analyzed. The composition of lava from fissure 13 became less evolved with subsequent reactivations and, by 24 May, was indistinguishable from the composition of other phase 2 lava.

Other fissures reactivated during late June (fissure 18) and early July (fissure 22). Samples

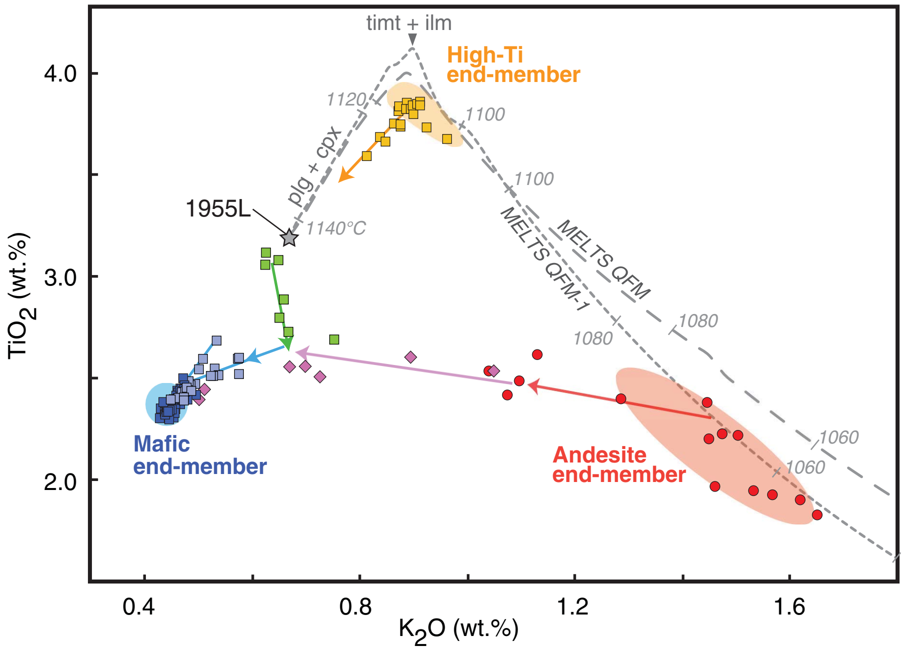


Fig. 4. K_2O versus TiO_2 plot of the 2018 eruption samples with possible mixing, end-members, and fractionation paths. Whole-rock EDXRF data [from data S1 (*7*)]. Symbols as in Fig. 2; colored regions distinguish inferred mixing end-members. Colored arrows show the direction of compositional change with time for each stage. Dashed lines are two MELTS (*12*, *27*) fractional crystallization models performed using the bulk composition of late 1955 samples (*24*) as starting composition, at fO_2 conditions along the QFM and QFM-1 buffers. Temperature steps are shown and compare favorably with our calculated magma temperatures (Table 1).

were not collected until the eruption stopped, because the vents were inaccessible after the fissure 8 flow began, so their exact timing is unknown. Near-vent samples are given a maximum formation date of 12 July 2018. These samples all show anomalously evolved compositions compared to fissure 8 lavas from the same period (Figs. 2 to 4).

This change with reactivation was most pronounced at fissure 22, which produced audible explosions and built a cinder cone over the vent. The composition of vent spatter overlaps with that of the late-erupted fissure 17 lava (Figs. 2 to 4) and was thus quite viscous (table S1). Glass SiO_2 averaged 55.6 wt %, and MgO 3.0 wt %. Phenocryst compositions varied widely, overlapping with crystals from fissure 8 (Fo_{78} olivine) and fissure 17 (high-Fe orthopyroxenes, An_{50} plagioclase) lavas (Fig. 4 and fig. S1). Olivine compositions showed a an exceptionally wide range, from Fo_{63} to Fo_{80} (Fig. 5). Physical mingling of two distinct melts is visible in SEM images of fissure 22 spatter (fig. S2).

Interpreting magmatic processes

The 2018 Kīlauea eruption recorded a complex story of magma storage, mixing, and migration. During the past two-plus centuries, magma has repeatedly been injected into the LERZ (1790, 1840, 1924, 1955, and 1960; fig. S3), mixing with differentiated magmas from prior intrusions (*1*). Geochemical evidence suggests that at least two separate LERZ magma bodies interacted and mixed with hotter, less fractionated summit or deep rift zone magma during the 2018 eruption (Fig. 6). Early phase 1 lavas constitute one source (the “High-Ti” end-member of Fig. 4) and fissure 17 lavas a second stored magma source (the “Andesite” end-member). Phase 3 lavas represent another end-member, which appears to source from an olivine-controlled, tholeiitic basalt magma with similarities to both the Halemaʻumaʻu and Puʻu ʻŌʻō lavas (“Mafic” end-member).

Differentiated and stored magmas

Whole-rock and mineral compositions of initial phase 1 lavas (“High-Ti” end-member) form distinct clusters and showed little to no overlap with other erupted lavas or Puʻu ʻŌʻō and summit lavas (Figs. 2 to 5 and fig. S1). The evolved composition, low temperature, and presence of orthopyroxene indicate early phase 1 differentiated from magma previously trapped in the rift zone [e.g., (*1*)] that was forced to the surface by the intruding dike (Fig. 3). Most early phase 1 lavas could have formed by differentiation of late 1955 magma (Fig. 4), on the basis of MELTS calculations (*12*). The rare presence of ilmenite agrees with this model, which shows the basalts to be at maximum iron enrichment and onset of Fe-Ti oxide crystallization. Several early phase 1 samples were either less differentiated or lie along a possible mixing line with later phase 1 compositions (Fig. 4). Late phase 1 lavas were too mafic to be differentiated from 1955 lava (Fig. 4), and mineralogical and compositional variations indicate that they were likely mixtures between the three magmatic end-members.

The fissure 17 andesite flows had a broad range of orthopyroxene Mg compositions from enstatite 80 to 45 (fig. S1). The distinctive mineralogy and chemistry suggested the andesite formed as an isolated magma body rather than by mixing with other magmas. We considered that the andesite resulted from mixing between a dacite body encountered during drilling in 2005 by Puna Geothermal Venture about 2 km away (*13*) and one of the other end-members. However, the andesite does not sit on any reasonable mixing lines with the dacite (data S2 and fig. S4).

Late 1955 magma is also a possible source for the andesite, depending on cooling history and oxygen fugacity (fO_2) conditions (Fig. 4). It cannot be the source of both the andesite and the early phase 1 lava, as the compositional gap is too great and the andesite volume much larger. We could not obtain reasonable matches between MELTS differentiation paths and the andesite using other characterized local magma sources in the LERZ (1960 or 1790) as starting compositions (Fig. 4). A specific source may not be identifiable if larger, long-lived rift magma bodies were amalgamations of repeated dike injections over centuries within the lower east rift zone (*14*, *15*).

Mixing and hybridization of lavas

Lavas erupted during late phase 1 and phase 2 appeared to be part of a continuum of mixing between early phase 1 and phase 3 compositions (Figs. 2, 3, and 5). Simple mixing and hybridization would appear as a straight line between the High-Ti and Mafic end-members (Fig. 4) if it were caused by a mafic dike intruding a single differentiated rift stored magma body (*1*, *2*).

Many late phase 1 and early phase 2 lavas showed unexpected signs of mixing with the fissure 17 andesite (Figs. 4 and 6). Lava collected from vents 20, 19, and 13 on 15 to 18 May, just prior to the onset of high-volume phase 2 eruptions, showed increasing mixing with the andesite, despite being up to 2.5 km from fissure 17 (Figs. 1 and 4 and data S1). The final four samples of andesite erupted on 20 to 22 May had about 25% lower K_2O, indicating mixing with mafic magma from the main dike system (Fig. 4). Vents 18 and 22 reactivated during phase 3, and their compositions also lie on a mixing line with the andesite

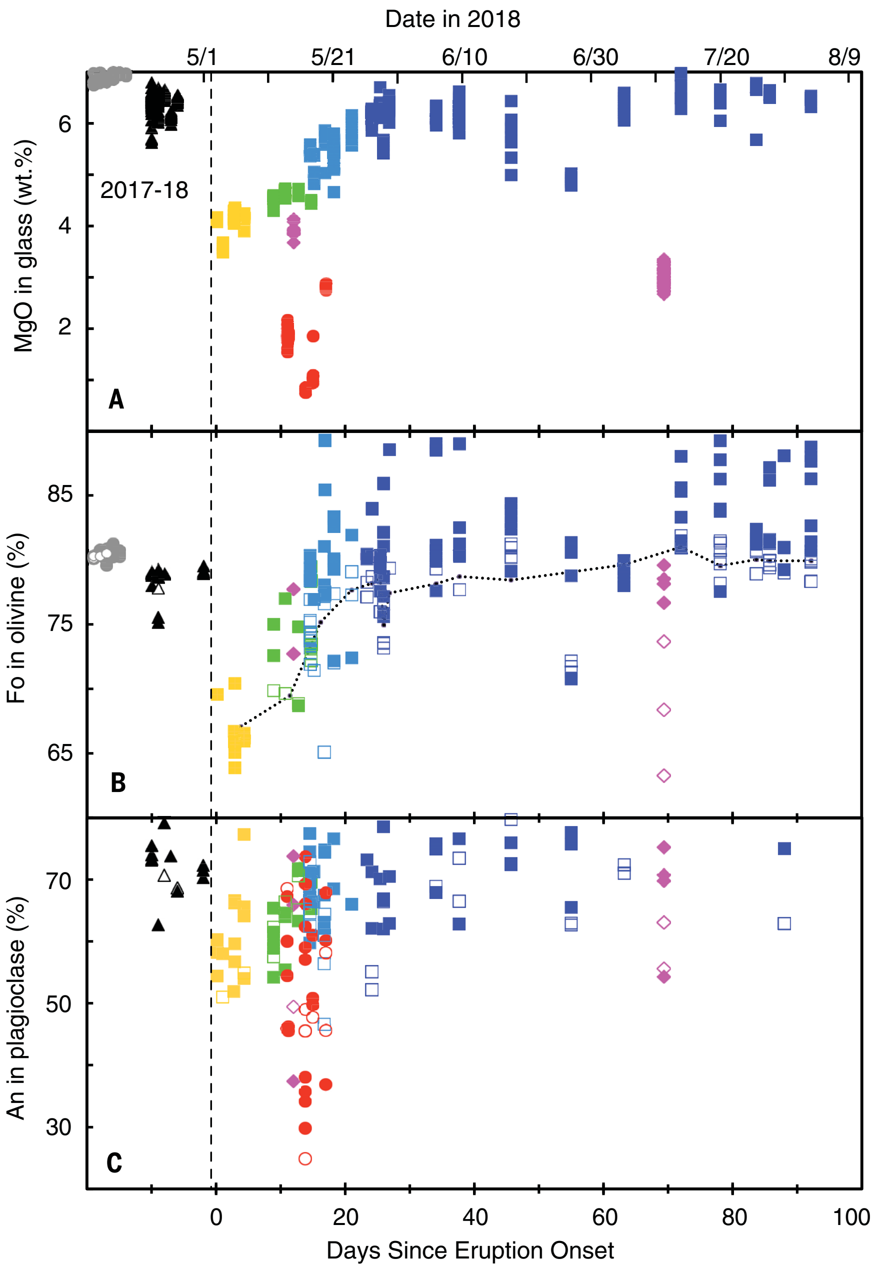


Fig. 5. Temporal variation in glass and mineral compositions. Plot of (**A**) MgO in matrix glass, (**B**) Fo mol % in olivine, and (**C**) An mol % in plagioclase during the 2018 eruption. To left of dashed line are compositions from Pu'u 'Ō'ō and Halema'uma'u lake lava from 2017–2018, including a breakout on Pu'u 'Ō'ō 2 days before the LERZ eruption began. Symbols are as in Fig. 2, with the addition of gray circles for summit samples. For the phenocryst analyses, open symbols are rims, solid are mineral cores. Dotted line in (B) is approximate Fo in equilibrium with maximum glass MgO in (A) using calculations of (*16*); note correlation with rim compositions. Glass MgO is very low in the initial phase 1 lava, increases until about 26 May, then levels out at approximately Pu'u 'Ō'ō levels, with the notable exception of fissure 17 and the anomalous reactivated fissures. One fissure 8 flow sample at day 55 is also anomalously low in glass and olivine MgO, but it was collected 12 km from the vent and is crystal-rich; other samples were collected within a few kilometers of their vents. Data from EMP spot-analyses (data S3 and S4) (*7*).

(Fig. 4 and data S1). One fissure 22 sample contained low-Mg orthopyroxenes (fig. S1), and Fe-Ti oxides found only in the andesite, and showed mingling between basalt and andesite glasses (fig. S2). The complex mixing patterns indicate that fissure 17 must be an en echelon fissure that intersected the andesite body but was also connected at depth to the main dike. The areal distribution of mixing (Fig. 6) suggests that the fissure 17 dike system may extend at depth several kilometers to the southwest of the surface vent (Fig. 1) and that the andesite body could be larger than the erupted volume.

The arrival of clearly identifiable mafic magma in the intruding dike began on 17 to 18 May at the start of phase 2, marked by an abrupt rise in lava temperature, a shift in composition, rapidly increasing effusion rates, and the diminishing of local deformation and seismicity (Figs. 2, 3, and 5). Both new and reactivated fissures along the main trend showed rapid compositional fluctuations, as more primitive magma mixed with and flushed out the stored magma by 24 May. Extremely Mg-rich olivine crystals (Fo_{88-89}) first appeared on 20 May. The Fo_{88-89} olivine cores would have formed at temperatures of 1270° to 1290°C in magma with ~13 to 14 wt % MgO (*8*, *16*), conditions that generally only occur in the summit magma reservoir or possibly in the deep rift. The outer rims of the olivine crystals were around Fo_{75}, generally in equilibrium with the erupted lavas (Fig. 5B). Kink banding was observed in a few larger (>2 mm) olivines during phase 2, implying minor entrainment from a cumulate source, likely from a deeper part of the rift zone (*17*).

Voluminous mafic lava

Phase 3 lavas erupted from fissure 8 were reasonably homogeneous basalts similar in bulk composition to recent Pu'u 'Ō'ō or Halema'uma'u lava, but with distinctly elevated incompatible-element concentrations (K_2O, La, Ti, Zr) (Table 1 and Figs. 2, 4, and 7). As phase 3 progressed, MgO in glass shifted to higher values (from ~6.1 wt % to 6.6 wt %), and the mineral phenocryst assemblage gradually changed from olivine, clinopyroxene, and plagioclase to nearly olivine-only during the last month. Olivines in early phase 3 continued to show the broad range of core compositions (Fo_{77-89}) that were first observed in phase 2. Most olivine rims (Fo_{77-80}) were in equilibrium with the host melts (Fig. 5B) as they became more MgO-rich over time. The persistent high-Mg olivines (Fo_{88-89}) that we found in phase 3 lavas contained Cr-spinel, were normally zoned, and had euhedral forms, characteristics of a magmatic rather than rift cumulate origin. Olivines with Fo_{89} compositions are extremely rare in subaerial Kīlauea magmas (*11*, *18*), but have been found in the submarine portion of the East Rift zone (*17*), transported from either the summit magma reservoir or the deep rift zone.

Kīlauea magma compositions and summit storage

Magma ascends from the mantle into the summit storage system of Kīlauea with a recharge rate of at least 0.1 km^3/year (*19*). The system was composed of a shallow (1- to 2-km depth),

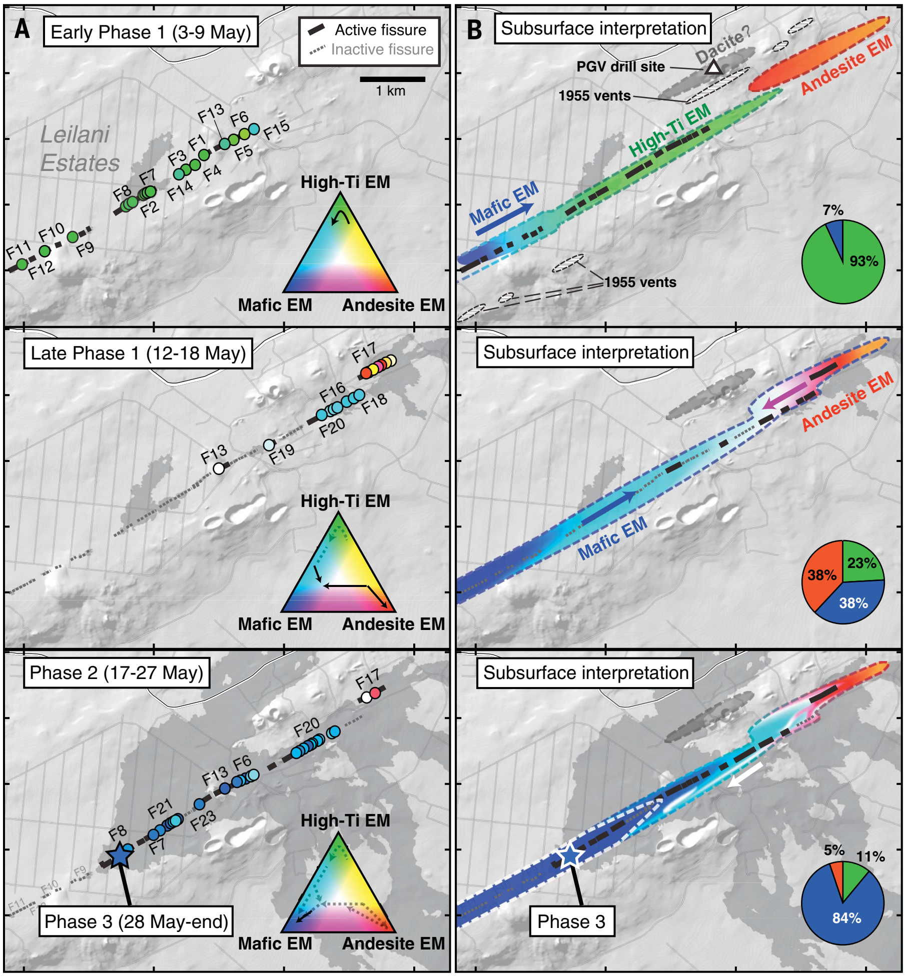


Fig. 6. Evolution of magma compositions with spatial location. (**A**) Series of panels showing samples repositioned along their active fissure of origin (colored circles) for each of the four phases (early and late phase 1, 2, 3), where the along-fissure position is approximated. A red-green-blue color mixing scheme is used to represent mixing of the three main end-member magmas. Note that the High-Ti end-member magma is green here (gold in other figures). RGB triangle shows color scheme and inferred mixing paths for the different periods. (**B**) Interpretative magma location maps corresponding to the same four phases. Colored regions represent the possible extent of magma underneath the surface. Colors show extents of mixing between end-members. Only general trends are depicted, as complex compositional variations occurred even within each of phases 1 and 2. Pie charts show percentages of each magma end-member erupted for each phase calculated from the TiO_2-K_2O relationships.

small magma chamber under Halemaʻumaʻu connected to a deeper (3 to 5 km), larger (3 to 20 km^3) chamber beneath the southeastern region of Kīlauea caldera with possible additional storage in a deep rift system (*19*, *20*). From 2008 to 2018, lava circulated, cooled, and degassed within the shallow Halemaʻumaʻu lava lake prior to passing back through the deeper chamber and out through a shallow dike (2- to 3-km depth) feeding the Puʻu ʻŌʻō vent (*11*, *19*). The volume of the 2018 summit collapse suggests that ~1 km^3 of magma was drained from the shallow Halemaʻumaʻu body (*5*).

Isotopic (Sr, Nd, Pb) and incompatible-element (K, Ti, Zr, La, Nb) compositions of olivine-controlled Kīlauea summit magmas vary over time and are thought to reflect changes in mantle source compositions (*21*). Distinct changes in incompatible-element concentrations of olivine-fractionated summit and rift lavas have been previously attributed to (i) storage of small, discrete magma "batches" in a plexus of dikes and sills with residence times of one to two decades (*18*); (ii) nearly continual mantle recharge of a small summit magma chamber (<0.5 km^3) with a residence time of <10 years (*21*); and (iii) a compositionally and thermally zoned magma chamber with variable vertical mixing (*17*) and much longer residence times.

Incompatible-element concentrations of K_2O and TiO_2 (Fig. 7) in olivine-fractionated Puʻu ʻŌʻō lavas (>6.8% MgO) gradually declined from 1985 to 2000, remained relatively flat through 2015, and began increasing systematically from 2016 to 2018. The correlation of Halemaʻumaʻu and Puʻu ʻŌʻō K_2O and TiO_2 values from 2010 to 2018 (Fig. 7) supports the genetic linkage between the two vents (*11*). The gradual change in composition coupled with the large erupted volume (~3.5 km^3) from 1983 to 2018 cannot be explained by the "dikes and sills" model but is compatible with the other two models. K_2O, TiO_2, Zr, and other incompatible elements were higher at the start of phase 3 than for the past 2 years (Figs. 2 and 7) and remained relatively constant for the remainder of the 2018 eruption. The small magma chamber model (<0.5 km^3) cannot easily explain the appearance of nearly 1 km^3 of "new" magma and disappearance of an equal volume of "Halemaʻumaʻu" magma. A more likely model is that denser, degassed magma draining from the shallow Halemaʻumaʻu chamber vertically mixed with hotter magma bearing a higher incompatible signature (K_2O, TiO_2, Zr,) residing either deeper in the summit chamber or the deep rift zone.

Magma transport

Early phase 3 lavas had olivine cores with distribution peaks at Fo_{78-79} (similar to 2017–2018 Puʻu ʻŌʻō cores), at Fo_{80-81} (similar to 2017–2018 Halemaʻumaʻu cores), and at Fo_{88-89} (data S4 and fig. S5), representative of a higher-temperature component. This suggests that cooler magmas similar to Halemaʻumaʻu and Puʻu ʻŌʻō magmas were mixing with deeper, hotter lava in the summit chamber or deep rift to form the dike magma. During the last 20 days of the 2018 eruption, olivine cores show bimodal peaks at Fo_{80-81} and Fo_{88-89} (data S4 and fig. S5) with only a few Fo_{78-79} cores. The combination of the distribution of olivine compositions, the absence of phenocrysts other than olivine, and the higher glass MgO compositions are highly suggestive that lavas erupted during the last 20 days were derived by mixing of shallow (cooler with lower K_2O) and deep (hotter with higher K_2O) components from the summit.

The initial dike propagated downrift from Puʻu ʻŌʻō on 30 April 2018 and erupted on 3 May. By 24 May, the mafic magma reached a stable, olivine-controlled composition with elevated incompatible-element concentrations relative to Puʻu ʻŌʻō and Halemaʻumaʻu lavas (Figs. 2, 5, and 7). The 3- to 4-week arrival time for mafic magma is consistent with intervals documented in 1955 and 1960. However, the final volume of mafic lava erupted in 2018 was about 10 times larger than the previous eruptions (*1*, *2*, *18*) and appears to have flushed the system of differentiated magma. It is unclear if magma was transported from

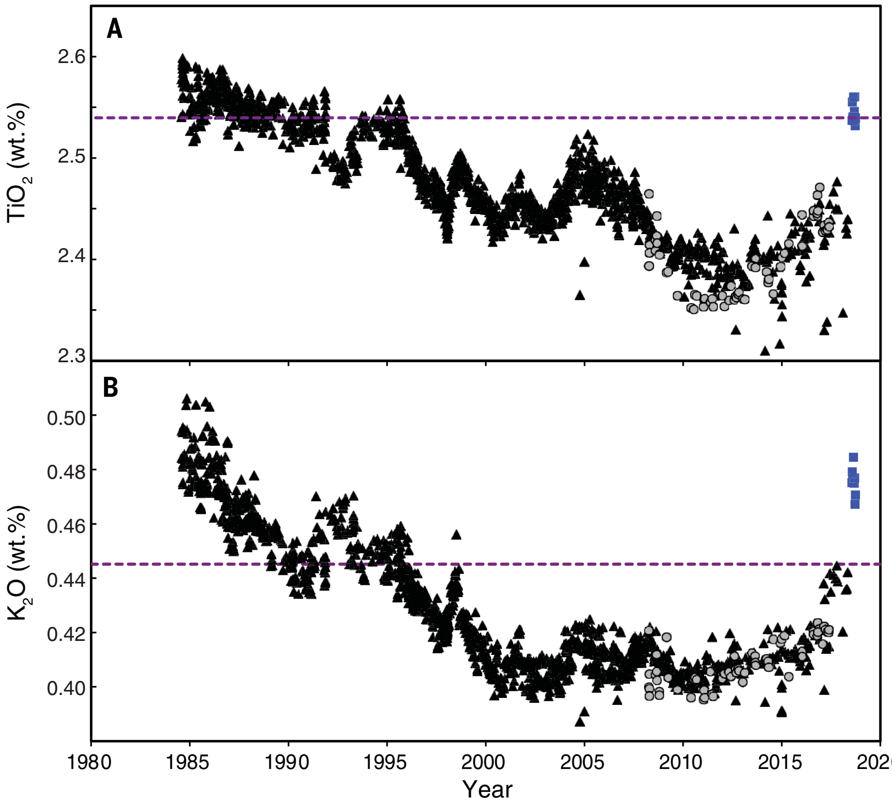


Fig. 7. Temporal variation of TiO_2 and K_2O during 1983–2018 Kīlauea eruptions. Whole-rock (**A**) TiO_2 and (**B**) K_2O (both normalized to 7 wt % MgO) are plotted versus time. Obviously evolved magmas (less than 6.8 wt % MgO) are excluded. Black triangles are Puʻu ʻŌʻō, gray circles are Halemaʻumaʻu lava lake, and blue squares are 2018 LERZ fissure 8 samples; dashed line marks the average composition of the April and September 1982 summit eruptions [data from (28, 29)]. TiO_2 and K_2O by WDXRF; from (7, 11) (data S2).

the summit to the eruption site using the established Puʻu ʻŌʻō dike system at about 2- to 3-km depth (22) or if transport switched to the deep rift plumbing system (19, 20). The most magnesian olivines from the Puʻu ʻŌʻō eruption were Fo_{87} (23), suggesting that the abundant $Fo_{88\text{-}89}$ olivines could not have been entrained by transport through the shallow Puʻu ʻŌʻō dike system. The $Fo_{88\text{-}89}$ olivines may have originated from deeper parts of the rift system, picked up during transport (17). Alternatively, magma carrying the high-Fo olivines could have exited from a deeper level of the summit chamber directly into the deep rift system (17) (fig. S3).

Petrology as a volcano-monitoring tool

The use of near-real-time compositional data was incorporated into the eruption monitoring and response. Geochemical results were reported to HVO staff and posted on the communications platform for field crews when the data seemed relevant for response purposes. Early phase 1 lava was cool and contained abundant plagioclase microlites, indicating highly viscous lava (approximate bulk magma viscosity 6600 Pa·s, table S1) that was likely to behave as slow-moving flows. Late phase 1 lava was 10° to 15°C hotter and had fewer microlites. The viscosity (2900 Pa·s) was lower, but the lava was still sluggish, and we anticipated the arrival of hotter and less viscous magma from Puʻu ʻŌʻō or the summit reservoir. The lava chemistry gave us early indication of hotter lava on 13 May, and we alerted science and field teams to the change. The 17 to 18 May arrival of less evolved and much hotter magma was indicated by bulk-rock MgO and CaO reaching Puʻu ʻŌʻō compositions. Field crews were again alerted that the substantial compositional change was underway and could lead to greater effusion of hotter, more fluid lava. Phase 2 and 3 lavas (~1140° to 1150°C) had viscosities in the typical range for ʻaʻā and pāhoehoe (~1150 Pa·s), increasing the likelihood of faster-moving flows. The chemical shifts correlated well with deformation and seismic signals recorded on nearby stations, showing the expansion of the dike (northward movement) essentially leveling out around that same time (Fig. 3).

A similar, if less extreme, eruptive sequence occurred during the 1955 LERZ eruption (24), implying that future rift zone eruptions may start deceptively small as older, stored magma erupts. Once the magma pathway opens, and fresher, hotter magma arrives, rift zone eruptions can rapidly switch to large, fast-moving lava flows.

The explosive nature of fissure 17 lava was also consistent with its unusually evolved chemistry. The viscosities we calculated were orders of magnitude higher than for other units (up to ~2 × 10^6 Pa·s).

A critical part of a volcano response is outreach and communications with the public. HVO's communications included geochemical information in press releases, interviews, and social media posts, and found an audience surprisingly interested in the seemingly esoteric questions of "new" and "old" magma sources and transport within Kīlauea.

Conclusion

The 2018 eruption of Kīlauea Volcano provided an opportunity to test a rapid-response geochemical analysis routine developed during the continuous eruption at Puʻu ʻŌʻō. The effort yielded critical information for hazards assessment and risk mitigation during the eruption. The collection of this large suite of lava samples and rich geochemical data set also allowed estimates of magma composition, mixing, temperature, viscosity, and travel time down the rift zone. Notably, the rapid-response data also proved highly suitable for geochemical modeling. The extremely large sample set that was made possible by our strategy filled in many gaps and allowed us to construct a more complete picture of the complex lava interactions. Based on the success of the HVO-UH Hilo geochemistry monitoring collaboration, other volcano observatories may benefit from similar efforts.

Materials and methods

USGS-HVO field crews collected 113 molten or recently solidified samples during the eruption. Samples were delivered 2 to 12 hours after collection to UH Hilo, where they were dried, given a quick petrographic overview, powdered, and pressed into pellets for analysis. We analyzed all samples by EDXRF for a limited suite of major (Ca, K, Ti, Mg) and trace (Rb, Sr, Zr, Y Nb) elements in the whole rock. The turnaround time from rock to data was 1 to 2 hours, and we analyzed most within 24 hours of field collection. We analyzed a subset of samples during and after the eruption by conventional WDXRF spectroscopy for a full suite of elements. CaO and MgO concentrations were used as geothermometers to estimate magma temperatures. We also determined matrix glass and phenocryst compositions by EMPA on polished thin sections and grain mounts from representative samples.

REFERENCES AND NOTES

1. T. L. Wright, R. S. Fiske, Origin of the differentiated and hybrid lavas of Kilauea Volcano, Hawaii. *J. Petrol.* **12**, 1–65 (1971). doi: 10.1093/petrology/12.1.1
2. R. T. Helz, T. L. Wright, Differentiation and magma mixing on Kilauea's east rift zone: A further look at the eruptions of 1955 and 1960. Part I. The late 1955 lavas. *Bull. Volcanol.* **54**, 361–384 (1992). doi: 10.1007/BF00312319
3. M. O. Garcia, R. A. Ho, J. M. Rhodes, E. W. Wolfe, Petrologic constraints on rift-zone processes: Results from episode 1 of the Pu'u 'O'o eruption of Kīlauea volcano, Hawai'i. *Bull. Volcanol.* **52**, 81–96 (1989). doi: 10.1007/BF00301548
4. C. R. Thornber, "Magma-reservoir processes revealed by geochemistry of the Pu'u 'Ō'ō - Kūpaianaha Eruption" in *The Pu'u 'Ō'ō – Kūpaianaha Eruption of Kīlauea Volcano, Hawai'i: The First 20 Years*, C. Heliker, D. A. Swanson, T. J. Takahashi, Eds., U.S. Geol. Survey Prof. Paper **1676**, 121–136 (2003).
5. C. A. Neal *et al*., The 2018 rift eruption and summit collapse of Kīlauea Volcano. *Science* **363**, 367–374 (2019). doi: 10.1126/science.aav7046; pmid: 30538164
6. Materials and methods are available as supplementary materials
7. R. L. Lee *et al*., Whole-rock and glass chemistry of lava samples collected during the 2018 Lower East Rift Zone eruption of Kilauea. U.S. Geol. Surv. Data Release, https://doi.org/10.5066/P9LVY7GV (2019).
8. R. T. Helz, C. R. Thornber, Geothermometry of Kilauea Iki lava lake, Hawaii. *Bull. Volcanol.* **49**, 651–668 (1987). doi: 10.1007/BF01080357
9. R. T. Helz, N. G. Banks, C. Heliker, C. A. Neal, E. W. Wolfe, Comparative geothermometry of recent Hawaiian eruptions. *J. Geophys. Res.* **100** (B9), 17637–17657 (1995). doi: 10.1029/95JB01309
10. E. W. Wolfe, J. Morris, Sample data for the geologic map of the island of Hawaii. *USGS Miscellaneous Investigations Series* **I-2524-B**, scale 1:100,000. https://doi.org/10.3133/i2524B (1996).
11. C. R. Thornber, T. R. Orr, C. Heliker, R. P. Hoblitt, "Petrologic testament to changes in shallow magma storage and transport during 30+ years of recharge and eruption at Kīlauea Volcano, Hawai'i" in *Hawaiian Volcanoes: From Source to Surface*, R. J. Carey, V. Cayol, M. P. Poland, D. Weis, Eds. Am. Geophys. Union, Geophys. Monograph **208**, 147–188 (2015).
12. M. S. Ghiorso, R. O. Sack, Chemical mass transfer in magmatic processes. IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contrib. Mineral. Petrol.* **119**, 197–212 (1995). doi: 10.1007/BF00307281
13. W. Teplow *et al*., Dacite melt at the Puna Geothermal Venture Wellfield, Big Island of Hawaii. *GRC Trans.* **33**, 989–994 (2009).
14. R. B. Moore, Distribution of differentiated tholeiitic basalts on the lower east rift zone of Kilauea Volcano, Hawaii. *Geology* **11**, 136–140 (1983). doi: 10.1130/0091-7613(1983)11<136:DODTBO>2.0.CO;2
15. K. M. Cooper, M. R. Reid, M. T. Murrell, D. A. Clague, Crystal and magma residence at Kilauea Volcano, Hawaii: ^{230}Th-^{226}Ra dating of the 1955 east rift eruption. *Earth Planet. Sci. Lett.* **184**, 703–718 (2001). doi: 10.1016/S0012-821X(00)00341-1
16. P. L. Roeder, R. F. Emslie, Olivine-liquid equilibrium. *Contrib. Mineral. Petrol.* **29**, 275–289 (1970). doi: 10.1007/BF00371276
17. D. A. Clague, J. G. Moore, J. E. Dixon, W. B. Friesen, Petrology of submarine lavas from Kilauea's Puna Ridge, Hawaii. *J. Petrol.* **36**, 299–349 (1995). doi: 10.1093/petrology/36.2.299
18. R. T. Helz, D. A. Clague, T. W. Sisson, C. R. Thornber, "Petrologic Insights into Basaltic Volcanism at Historically Active Hawaiian Volcanoes" in *Characteristics of Hawaiian Volcanoes*. *U.S. Geol. Surv. Prof. Pap.* **1801**, 237–262 (2014).
19. M. P. Poland, A. Miklius, E. K. Montgomery-Brown, "Magma supply, storage, and transport at shield-stage Hawaiian volcanoes" in *Characteristics of Hawaiian Volcanoes*. *U.S. Geol. Surv. Prof. Pap.* **1801**, 179–234 (2014).
20. P. T. Delaney, R. S. Fiske, A. Miklius, A. T. Okamura, M. K. Sako, Deep magma body beneath the summit and rift zones of Kilauea Volcano, Hawaii. *Science* **247**, 1311–1316 (1990). doi: 10.1126/science.247.4948.1311; pmid: 17843792
21. A. J. Pietruszka, D. E. Heaton, J. P. Marske, M. O. Garcia, Two magma bodies beneath the summit of Kīlauea Volcano unveiled by isotopically distinct melt deliveries from the mantle. *Earth Planet. Sci. Lett.* **413**, 90–100 (2015). doi: 10.1016/j.epsl.2014.12.040
22. A. T. Okamura, J. J. Dvorak, R. Y. Koyanagi, W. R. Tanigawa, "Surface deformation during dike propagation" in *The Pu'u 'Ō'ō Eruption of Kīlauea Volcano, Hawai'i: Episodes 1 through 20, January 3, 1983 through June 8, 1984*, E. W. Wolfe, Ed., U.S. Geol. Survey Prof. Paper **1463**, 165-180 (1988).
23. M. O. Garcia, J. M. Rhodes, E. W. Wolfe, G. E. Ulrich, R. A. Ho, Petrology of lavas from episodes 2-47 of the Pu'u 'Ō'ō eruption of Kīlauea Volcano, Hawai'i: Evaluation of magmatic processes. *Bull. Volcanol.* **55**, 1–16 (1992). doi: 10.1007/BF00301115
24. G. A. Mcdonald, J. P. Eaton, Hawaiian volcanoes during 1955. *U.S. Geol. Surv. Bull.* **1171** (1964).
25. Hawaiian Volcano Observatory staff, Preliminary map of the 2018 lower East Rift Zone eruption of Kīlauea Volcano, Island of Hawai'i. U.S. Geological Survey, https://doi.org/10.5066/P9940GY8 (2018).
26. E. T. Endo, T. Murray, Real-time Seismic Amplitude Measurement (RSAM): A volcano monitoring and prediction tool. *Bull. Volcanol.* **53**, 533–545 (1991). doi: 10.1007/BF00298154
27. P. D. Asimow, M. S. Ghiorso, Algorithmic modifications extending MELTS to calculate subsolidus phase relations. *Am. Mineral.* **83**, 1127–1132 (1998). doi: 10.2138/am-1998-9-1022
28. M. O. Garcia, A. J. Pietruszka, J. M. Rhodes, A petrologic perspective of Kīlauea volcano's summit magma reservoir. *J. Petrol.* **44**, 2313–2339 (2003). doi: 10.1093/petrology/egg079
29. A. J. Pietruszka, M. O. Garcia, A rapid fluctuation in the mantle source and melting history of Kīlauea Volcano inferred from the geochemistry of its historical summit lavas (1790-1982). *J. Petrol.* **40**, 1321–1342 (1999). doi: 10.1093/petroj/40.8.1321

ACKNOWLEDGMENTS

Detailed comments by D. Clague, T. Wright, and an anonymous reviewer greatly improved the manuscript. UH Hilo students R. Sasaki, M. Warren, and K. Hutchison did much of the sample preparation and EDXRF work. USGS-HVO field crews collected the lava samples. WDXRF analyses were done by R. Conrey and L. Wagner. EMP analysis was done by D. Adams. Any use of brand names is for information only and does not imply endorsement by the U.S. Government. **Funding:** Geochemical work was funded by Cooperative Agreement G15AC00204 between USGS and the University of Hawaii. NSF grants BCS-0317528 and BCS 14-27950 to P. Mills and S.P.L. support the EDXRF lab at UH Hilo. NSF RAPID grant 1838502 supported T.S. **Author contributions:** C.G. oversaw data collection and analysis and conceptualization and writing of the manuscript. R.L.L. collected and managed geochemical data. T.S. contributed to data analysis and writing and produced several figures. S.P.L. contributed to EDXRF data collection. K.H. contributed to data analysis and writing. C.P. contributed to eruption chronology and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and sample metadata are available either in the supplementary materials or in USGS data release (*7*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaaz0147/suppl/DC1
Materials and Methods
Figs. S1 to S5
Table S1
Databases S1 to S4
References (*30*–*37*)

5 August 2019; accepted 8 November 2019
10.1126/science.aaz0147

RESEARCH ARTICLE SUMMARY

VOLCANOLOGY

Cyclic lava effusion during the 2018 eruption of Kīlauea Volcano

M. R. Patrick*, H. R. Dietterich, J. J. Lyons, A. K. Diefenbach, C. Parcheta, K. R. Anderson, A. Namiki, I. Sumita, B. Shiro, J. P. Kauahikaua

INTRODUCTION: The 2018 flank eruption and summit collapse of Kīlauea Volcano was one of the largest and most destructive volcanic events in Hawai'i in the past 200 years. The eruption occurred on the volcano's lower East Rift Zone (ERZ), draining magma at a high rate from the summit reservoir and triggering incremental collapse of the overlying caldera floor. Lava flows erupted for 3 months, destroying several residential subdivisions and burying miles of roads. The eruption rate exhibited cyclic behavior on multiple time scales, resulting in repeated lava breakouts and overflows. Multidisciplinary observations provide insight into the nature of these variations, driving forces in the magmatic system, and implications for hazard.

RATIONALE: Volumetric eruption rate is a primary control on the vigor and hazard of lava flows, but the processes that control its temporal variations are not well understood because of limited observational data. We integrated field observations, photos and video from time-lapse cameras and unmanned aircraft systems, seismic tremor, and infrasound to track the time scales and magnitude of fluctuations in eruption rate at the primary vent for the 2018 eruption on Kīlauea's LERZ. We combined these data with documentation of summit caldera collapses to investigate the origins and impacts of these fluctuations.

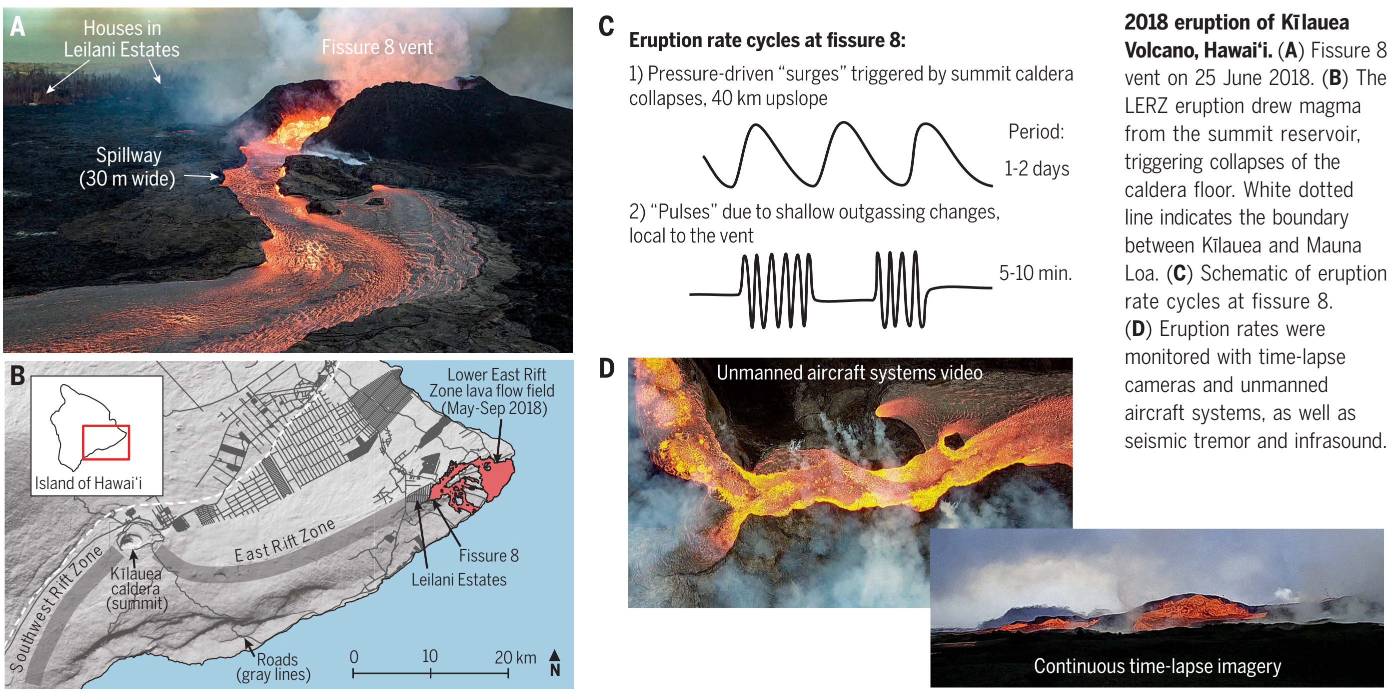
RESULTS: Cyclic variations in eruption rate occurred on two disparate time scales. First, short-term fluctuations ("pulses") in eruption vigor had periods of 5 to 10 min, but had no major implications for lava flow hazard. Flow rate in the lava channel was inversely related to fountaining and outgassing intensity at the vent. Second, long-term fluctuations ("surges") had periods of 1 to 2 days and began within minutes of episodic caldera collapse events at the summit, 40 km upslope. These surges triggered overflows from the channel that produced hazardous enlargement of the lava flow field, which could be forecast several hours in advance. We also show that seismic tremor and infrasound were correlated with lava flow eruption rates.

CONCLUSION: We conclude that the two types of eruption rate cycles were controlled by two distinct processes. Short-term pulses were driven by changes in outgassing efficiency of the lava at shallow depths. Long-term surges in eruption rate were driven by pressure transients induced by the summit collapses and transmitted through the magma conduit over a distance of 40 km. The pressure-driven surges in eruption rate demonstrate that the episodic rhythm of summit caldera collapse sequences may be imparted on the accompanying flank eruption. The surges also help to constrain the efficient hydraulic connection between Kīlauea's summit magma system and rift zones and demonstrate that pressure communication over distances of 40 km can occur on a time scale of minutes. Seismic tremor and infrasound may be effective proxies for lava flow eruption rates, allowing for improved tracking of lava flow hazards. Our multidisciplinary data provide a clear link between eruption rate fluctuations and their driving processes in the magmatic system. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay9070

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mpatrick@usgs.gov
Cite this article as M. R. Patrick *et al.*, *Science* **366**, eaay9070 (2019); DOI: 10.1126/science.aay9070



2018 eruption of Kīlauea Volcano, Hawai'i. (**A**) Fissure 8 vent on 25 June 2018. (**B**) The LERZ eruption drew magma from the summit reservoir, triggering collapses of the caldera floor. White dotted line indicates the boundary between Kīlauea and Mauna Loa. (**C**) Schematic of eruption rate cycles at fissure 8. (**D**) Eruption rates were monitored with time-lapse cameras and unmanned aircraft systems, as well as seismic tremor and infrasound.

CREDITS: USGS; UAS

RESEARCH ARTICLE

VOLCANOLOGY

Cyclic lava effusion during the 2018 eruption of Kīlauea Volcano

M. R. Patrick[1]*, H. R. Dietterich[2], J. J. Lyons[2], A. K. Diefenbach[3], C. Parcheta[1], K. R. Anderson[4], A. Namiki[5], I. Sumita[6], B. Shiro[1], J. P. Kauahikaua[1]

Lava flows present a recurring threat to communities on active volcanoes, and volumetric eruption rate is one of the primary factors controlling flow behavior and hazard. The time scales and driving forces of eruption rate variability, however, remain poorly understood. In 2018, a highly destructive eruption occurred on the lower flank of Kīlauea Volcano, Hawai'i, where the primary vent exhibited substantial cyclic eruption rates on both short (minutes) and long (tens of hours) time scales. We used multiparameter data to show that the short cycles were driven by shallow outgassing, whereas longer cycles were pressure-driven surges in magma supply triggered by summit caldera collapse events 40 kilometers upslope. The results provide a clear link between eruption rate fluctuations and their driving processes in the magmatic system.

Numerous communities have been destroyed or threatened by lava flows in recent decades (*1*–*3*), with recurring crises at volcanoes around the world, such as Nyiragongo (Democratic Republic of the Congo) (*4*), Piton de la Fournaise (Reunion Island) (*5*), and Etna (Sicily) (*6*, *7*). Another recent episode of destruction occurred with the 2014–2015 eruption of Fogo (Cape Verde), leaving ~1000 people homeless (*3*). At Kīlauea (Hawai'i), the 2018 eruption produced destructive lava flows (*8*) and the 2014–2015 crisis disrupted the lives of thousands of residents when lava stalled just short of destroying the town of Pāhoa (*9*). The Pāhoa crisis was just one episode of Kīlauea's long-lived Pu'u 'Ō'ō eruptions (1983 to 2018), which destroyed the town of Kalapana (*10*). Risk mitigation in such crises may include evacuation of residents, removal of property, relocation of critical infrastructure, or lava diversion in some cases (*1*). The success of this type of hazard response depends in large part on lava flow forecasting accuracy (*9*, *11*).

The volumetric eruption rate (effusion rate) is a primary factor controlling the advance rate, length, and coverage of lava flows (*12*, *13*). Effusion rate is an important input into quantitative models that can then be used to forecast advance rates and areal coverage (*14*, *15*). Most established relationships between effusion rate and flow behavior, however, are based on steady-state or time-averaged rates (*13*). Our understanding of effusion rate controls on flow behavior is challenged by large fluctuations in the rate (*16*). Furthermore, the time scales and driving forces of the effusion rate variability remain poorly understood (*13*). Determining whether these variations are deeply sourced (e.g., magma supply rate changes), shallowly rooted (e.g., outgassing or compositional changes), or result from surface processes (e.g., lava channel blockage) is often difficult (*17*–*19*).

Kīlauea Volcano has long been a focus for understanding lava flow behavior and hazards because of its history of sustained lava effusion (*20*). On 3 May 2018, eruptive activity began in the lower East Rift Zone (LERZ) (Fig. 1, A and B), ushering in the most destructive phase of volcanic activity in Hawai'i in the past 200 years (*8*). The main flow, erupted from fissure 8, was exceptionally well monitored (*8*, *21*). The robust observational dataset and accompanying geophysical signals captured cyclic fluctuations in lava eruption rate. This provided an opportunity to unravel the causative processes and their hazard implications.

The 2018 eruption of Kīlauea Volcano

Kīlauea erupted nearly continuously from 1983 to 2018 from vents on and near the Pu'u 'Ō'ō cone, on the volcano's middle ERZ (*10*, *22*, *23*). Lava flows, predominantly slow-moving, tube-fed pāhoehoe, covered 144 km^2 of land (Fig. 1A), with typical recent effusion rates of 2 to 6 $m^3\ s^{-1}$ (*24*, *25*). The eruption destroyed 215 structures (*23*). As the Pu'u 'Ō'ō eruption continued on the ERZ, a new vent opened at Kīlauea's summit in 2008 and persisted for the next 10 years, supplying a large, convecting lava lake (*26*, *27*). This joint activity marked the first time in the 200-year historical record that prolonged (>1 year) eruptions were concurrent on Kīlauea's summit and rift zone (*27*).

In March 2018, Kīlauea's magmatic system began to pressurize at a relatively high rate (*8*). Inflation was present at the summit and Pu'u 'Ō'ō, as well as along the 20-km-long ERZ conduit that connects these two eruption sites. Although similar previous episodes of inflation created new vents on or near Pu'u 'Ō'ō (*9*, *23*), the 2018 sequence culminated in an intrusion beginning on April 30 that propagated down-rift (east) from Pu'u 'Ō'ō into the volcano's LERZ (Fig. 1A) and terminated the 35-year eruption at Pu'u 'Ō'ō. The intrusion reached the surface and lava began erupting from new fissures in the Leilani Estates subdivision on May 3 (Fig. 1B). On May 27, activity focused on fissure 8, and lava advanced 13 km in 6 days to reach the ocean (Fig. 1, B and C). Fissure 8 continued as the dominant vent for the next 2 months. Preliminary estimates of effusion rate from the fissure 8 vent were in the range of 100 to 300 $m^3\ s^{-1}$ (dense-rock volume, with bubbles removed) (*28*), far surpassing the typical eruption rates of the previous several decades at Pu'u 'Ō'ō. By the end of major effusion in early August, the LERZ eruption had destroyed >700 structures, in addition to roadways and utility infrastructure.

The 2018 LERZ eruption was supplied by magma from Kīlauea's summit reservoir complex and middle ERZ (*8*). The summit lava lake, active for a decade in the Halema'uma'u pit crater, drained in early May, and the floor of Halema'uma'u began to collapse in a piecemeal manner. Beginning in late May and continuing into early August, broader parts of the caldera floor also began to collapse in large episodic events of several vertical meters in a piston-like manner, with recurrence intervals of 25 to 50 hours. Each collapse event released energy equivalent to a magnitude 5.3 earthquake (*8*). By early August, the caldera floor had subsided ~550 m (Fig. 1A).

Dual cycles of lava effusion

Within days of its onset in late May, the fissure 8 flow developed a stable proximal channel that persisted for the next 2 months. Low fountaining (20- to 80-m high) within the fissure 8 cone supplied lava to the proximal channel that consisted of a narrow cascading spillway 30-m wide and 300-m long (Fig. 1C and movie S1). The spillway emptied into a perched pāhoehoe channel up to 430-m wide.

The vigor of lava in the fissure 8 spillway displayed two time scales of cyclic fluctuation (Fig. 1C): short-term "pulses" had periods of 5 to 10 min and long-term "surges" seemed to occur soon after summit collapse events.

[1]U.S. Geological Survey, Hawaiian Volcano Observatory, Hilo, HI 96720, USA. [2]U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK 99508, USA. [3]U.S. Geological Survey, Cascades Volcano Observatory, Vancouver, WA 98683, USA. [4]U.S. Geological Survey, California Volcano Observatory, Menlo Park, CA 94025, USA. [5]School of Integrated Arts and Sciences, Hiroshima University, Higashi Hiroshima, Hiroshima 739-8521, Japan. [6]Graduate School of Natural Science and Technology, Kanazawa University, Kakuma, Kanazawa, 920-1192, Japan.
*Corresponding author. Email: mpatrick@usgs.gov

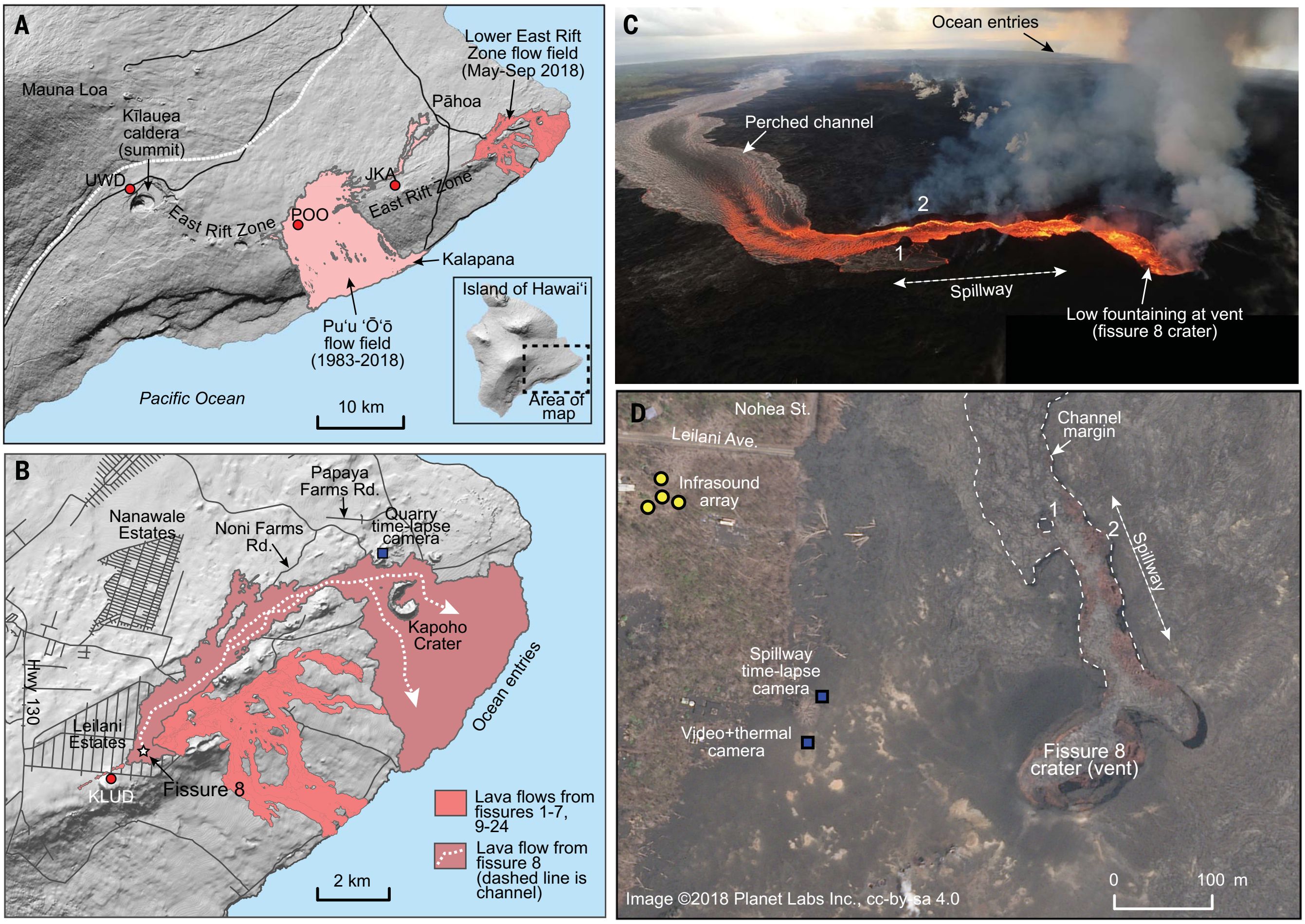


Fig. 1. Setting of the 2018 eruption. (**A**) Map of Kīlauea Volcano and the ERZ. Kīlauea forms the southeast portion of the Island of Hawaiʻi. The Puʻu ʻŌʻō eruption (1983 to 2018) ended at the onset of the LERZ eruption (May to September, 2018). Both eruptions were fed by magma supplied from the summit magma reservoir complex along a conduit that follows the ERZ. A large portion of the summit caldera floor collapsed and subsided in response to the LERZ eruption. White dotted line indicates the boundary between Kīlauea and Mauna Loa. (**B**) Close-up of the LERZ. Fissure 8 was active for ~2 months and formed the majority of the lava flow field. KLUD is a seismic station. (**C**) Aerial image looking east, showing the proximal sections of the fissure 8 flow. Lava flowed through a narrow spillway that then emptied into a broader perched channel. Numbers 1 (island) and 2 (high rim) mark spots shown in (D). Photo was taken July 29, 2018. (**D**) Post-eruption satellite image of the fissure 8 vent region and proximal lava channel. Measurements of lava level in the channel and flow velocities were targeted at the section of channel between points 1 and 2. Image courtesy of Planet Labs.

Pulses: short-term fluctuations

Numerous field crews reported substantial fluctuations in the level, speed, and agitation of lava in the spillway, occurring over time scales of minutes. Time-lapse imagery of the lava level and the seismic tremor amplitude [as shown by its proxy, real-time seismic amplitude measurement (RSAM) (*29*)] at nearby seismic station KLUD (Fig. 1B) provide details on this pulsing behavior and highlight the pulsing and nonpulsing regimes of vent activity (Figs. 2, A to C, and 3, A to C, and movie S2). During nonpulsing regimes, the lava level had a relatively steady height (Figs. 2A and 3B), with minor fluctuations (amplitudes of <2 m). RSAM was relatively steady (Fig. 3, A and B). During pulsing regimes, rapidly oscillating lava levels in the spillway were anticorrelated with seismic tremor (RSAM) and infrasound energy (Fig. 3C). Sporadic pulsing regimes comprised 35% of a 10-day period of observation in July (Fig. 3A), with durations of 1 to 18 hours (mean 6.4 hours).

During pulsing regimes, activity alternated between the lava channel and the fissure 8 crater. When the lava level in the channel peaked during pulsing, an unmanned aircraft system (UAS) and ground-based video showed a more rapid flow and a more agitated surface (less crust) (Figs. 2D and 4A and movie S3). Ground-based thermal images showed higher temperatures in the plume above the spillway (Fig. 3D), consistent with the visibly robust plume there and suggesting higher outgassing from the channel, whereas the gas plume at the vent was weaker and cooler, suggesting lower gas emission (Fig. 3D) and consistent with more subdued fountaining and weaker bubble bursts at the vent (Fig. 4A). At these times, RSAM and infrasound were relatively low (Fig. 3C).

Conditions were reversed during troughs in the pulsing. During these times, the lava in the channel was low and sluggish and the channel surface was more crusted and placid (Figs. 2D and 4B). Fountaining in the vent crater was more active, with extensive bubble bursting and a more robust gas plume (Fig. 4B), and thermal images indicated higher temperatures in the plume above the vent, suggesting higher outgassing there (Fig. 3D). RSAM and infrasound were high during these periods (Fig. 3C).

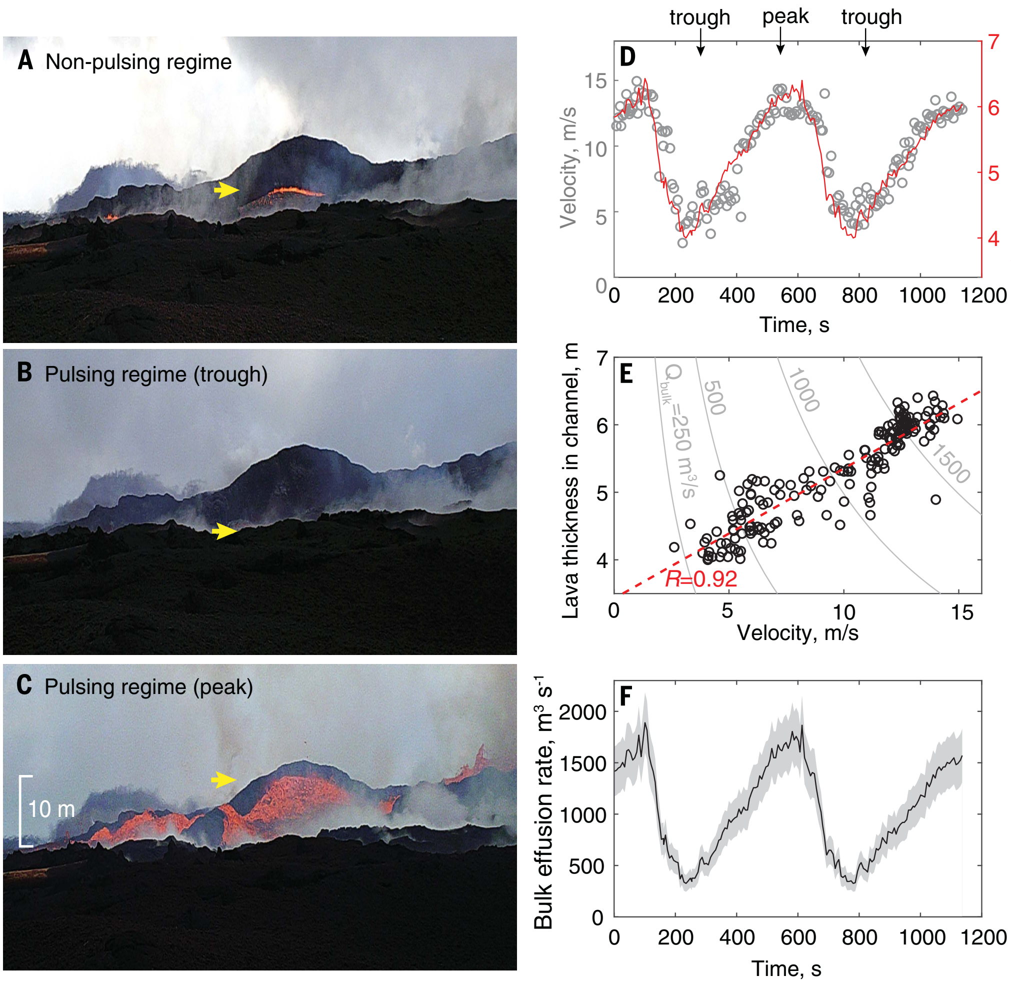


Fig. 2. Short-term cycles in effusion rate (pulses). (**A**) Typical lava level in the channel during nonpulsing eruptive behavior on 14 July 2018. (**B**) During the low levels (troughs) of pulsing, the level dropped several meters lower than normal, steady level. (**C**) During the high levels (peaks) of pulsing, the lava level rose several meters higher than the normal level. (**D**) Results from ground-based video of the lava channel spillway during pulsing regime on 19 July 2018. Velocity and lava level fluctuate in concert. (**E**) Correlation between velocity and lava level. (**F**) Time series of bulk effusion rates (i.e., not corrected for volume of bubbles) during pulsing. Gray area shows the uncertainty in effusion rate estimates based on ±1 m uncertainty in lava level in the channel. Panels (D) to (F) cover the time period of 8:10:06 to 8:29:52 HST on 19 July 2018.

Ground-based video of the spillway was used to quantify flow behavior and effusion rates. During a representative 20-min period in mid-July, two pulsing cycles were recorded, with velocities of 4 to 5 m s^{-1} during low lava levels and 12 to 15 m s^{-1} during high lava levels (Fig. 2D and movie S4). We observed a correlation (R = 0.92) between the height of the lava in the spillway and the velocity in that area (Fig. 2E). On the basis of these velocity and lava depth results and the measured channel width at this location (30 m), we estimated that bulk effusion rates decreased to ~350 m^3 s^{-1} in the troughs of the pulses and reached ~1700 m^3 s^{-1} at the peaks (Fig. 2F) [see Eq. 1 in (*30*)].

We used the relationship between lava level and velocity (Fig. 2E) to convert 4 hours of time-lapse images of lava level from July 14 (Fig. 3B) to bulk effusion rates (Fig. 5A). The effusion rate estimates had a general inverse relationship with RSAM, decreasing almost exponentially with higher RSAM values. Infrasound energy was also correlated inversely with bulk effusion rate during pulsing (Fig. 5B).

Surges: long-term fluctuations

Field crews reported that fissure 8 fountaining and flow in the spillway seemed to increase in vigor ("surge") after summit collapse events, which occurred with recurrence intervals of 25 to 50 hours. These visual observations were supported by changes in ground tilt, RSAM, and infrasound at LERZ stations that occurred within minutes of the summit collapse events. For instance, during July, LERZ tremor and infrasound energy exhibited conspicuous peaks that began within minutes of the summit collapse events (Fig. 6) (*30*). These RSAM and infrasound peaks suggested that eruptive vigor increased in the LERZ immediately after the summit collapses.

Lava-level changes in the spillway confirmed for us that there was a major increase in effusion rate at fissure 8 after summit collapse events. For example, on July 31, the lava level in the spillway was relatively low and steady in the 2 hours before the summit collapse event (Fig. 7 and movie S5). Within minutes, a clear rise in RSAM began, reaching a broad peak 2 to 3 hours after the event. The lava level in the area of the channel used for effusion rate measurements did not register a clear rise for ~30 min, but another, more sensitive, portion of the channel showed a rise in lava level that began within ~13 min of the summit collapse (fig. S13). Lava levels peaked ~3 hours after the summit collapse event. Using ground-based video to convert the lava level in the time-lapse images to effusion rate, the estimated bulk

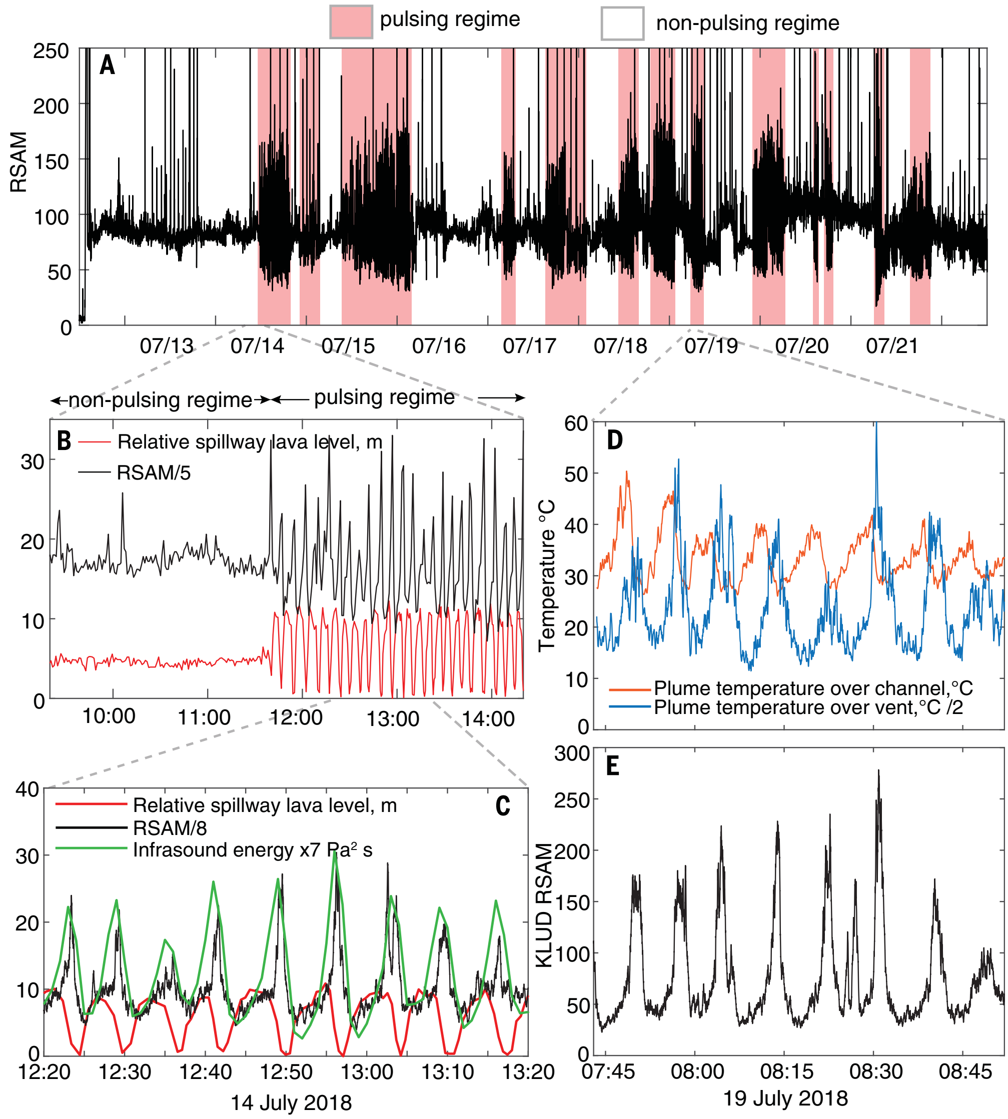


Fig. 3. Short-term cycles in eruption rates (pulses). (**A**) Ten days of RSAM showing nonpulsing behavior (white background) and pulsing behavior (pink background). Pulsing is evident by the higher variance. (**B**) Lava level and RSAM on 14 July 2018. Nonpulsing behavior is shown by a stable trend in these parameters, whereas pulsing behavior is distinctive with higher shared variance. (**C**) Lava level, RSAM, and infrasound energy on 14 July 2018, showing the inverse relationship between RSAM (and infrasound) and lava level. (**D**) Thermal image data of the pulses on 19 July 2018. Average temperatures are shown in small measurement windows in the plume above the lava channel spillway (orange line) and the vent where fountaining was occurring (blue line). (**E**) RSAM showing peaks that correlate with high vent temperatures (increased fountaining vigor).

effusion rates for July 31 were 300 to 500 $m^3 s^{-1}$ before the summit collapse event, with values peaking at ~1400 $m^3 s^{-1}$ ~3 hours after the event (Fig. 7F). Three other events in late July to early August illustrate this pattern of increased effusion rate after summit collapse events (Fig. 8). In these examples, precollapse bulk effusion rates were 300 to 700 $m^3 s^{-1}$, increasing to peaks of 1400 to 1700 $m^3 s^{-1}$ over ~4 hours. The precise onset times of lava-level rise in these three additional examples were obscured by natural variations, but we can constrain the onsets as occurring no later than 20 min after the summit collapses (fig. S13) (*30*). We chose these four surge events for analysis because of their good observational conditions; observation was limited in part by the period that the time-lapse camera was operating (*30*). Infrasound suggests that collapse-triggered surges in effusion rate were present as early as mid-June, if not earlier. However, infrasound also suggests that surging was absent or subdued for several weeks in early to mid-July despite continued summit collapses (*30*).

The four surge examples from late July and early August indicated that the effusion rates we estimated showed similar trends as the RSAM, with both having a broad peak after the summit collapse events with a prolonged (hours long) decline toward background values. The RSAM and bulk effusion rate (Fig. 5C) showed a linear correlation (R = 0.80). We disregard the July 26 surge because of the presence of intense pulsations around the time of the surge event. A linear correlation (R = 0.78) also existed between the bulk effusion rate and infrasound energy during the surges (Fig. 5D).

Time-lapse imagery in the distal channel 7.5 km from the vent recorded the downstream effects of these surges (movie S6). For example, on August 2, a summit collapse event at 11:55 Hawai'i Standard Time (HST) was followed ~20 min later by rising effusion rates at the fissure 8 vent. By 14:21 HST, a flood of lava was visible coming down the channel, triggering overflows (shown by white smoke

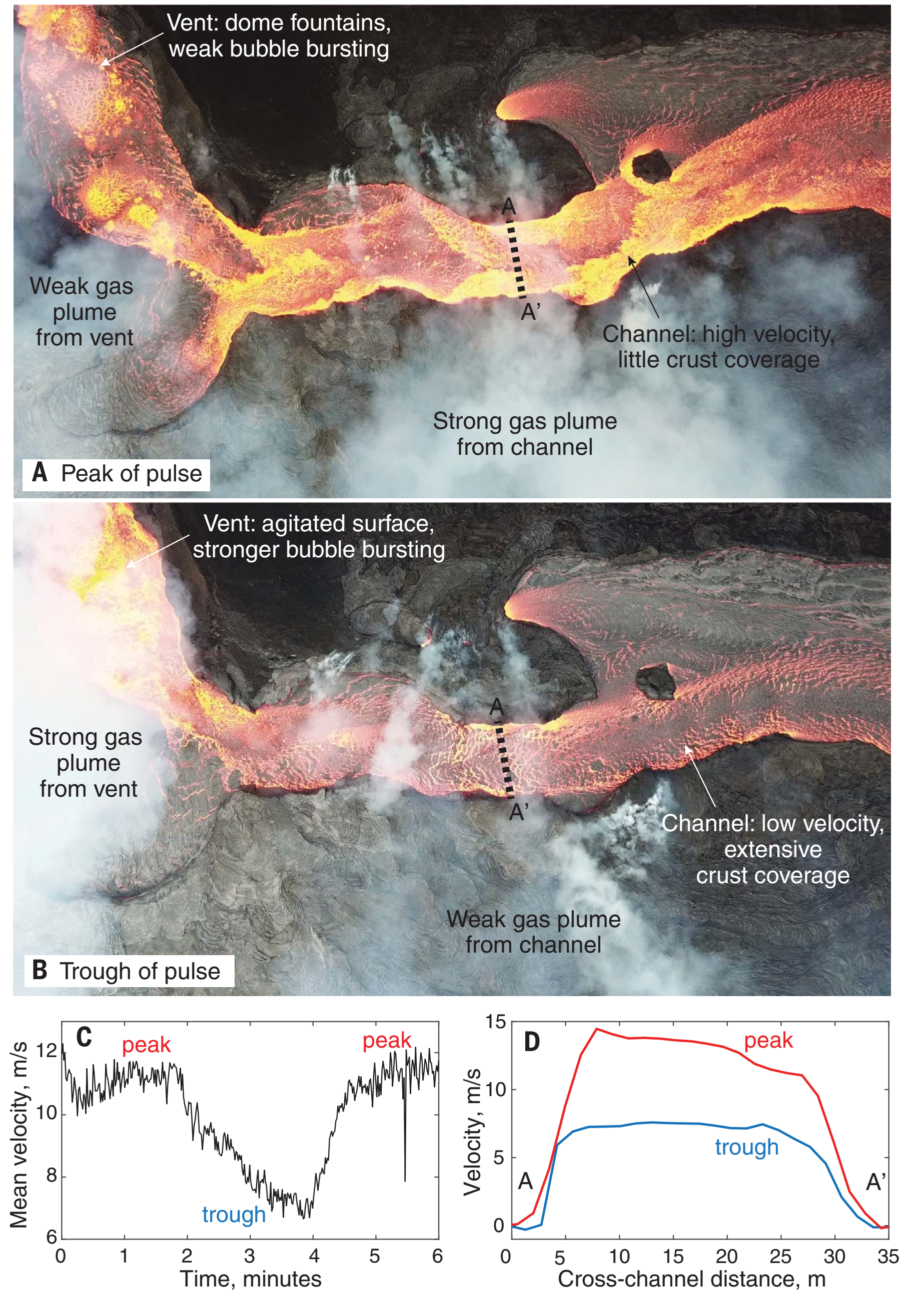


Fig. 4. UAS images of pulsing in the lava channel on 14 July 2018. (**A**) During the peak of pulses, the vent activity was subdued (low dome fountains and weak gas plume), but the upper channel (spillway) flow was faster and more vigorous and emitted a stronger plume. The vent pond dimensions were ~60 × 100 m. (**B**) During the trough of the pulses, the vent activity was heightened with vigorous bubble bursting and stronger gas plume, whereas the lava flowing in the channel was slower and more crusted. (**C**) One cycle of pulsing velocity from the UAS video. (**D**) Across-channel velocity profiles showing the change during a pulsing cycle.

from vegetation fires). Rising lava levels in the channel between 14:40 and 19:00 HST triggered inflation of the levees and "seeps" of spiny lava intruded through the levee (*31*).

Driving processes

What controls short-term fluctuations (pulses)?

The anticorrelation that we observed between bulk effusion rate and vent activity during pulsing behavior (Figs. 2F and 4) can be explained by variations in outgassing efficiency of lava at the vent. Stronger fountaining in the fissure 8 cone was associated with more efficient outgassing of lava at the vent, producing a denser, lower-volume lava flowing through the spillway (Fig. 4B). Weaker fountaining produced less efficient outgassing of lava, resulting in a bulkier, higher-volume, foamy lava pouring out of the crater into the spillway. This gas-charged lava then outgassed somewhat amidst the disruption of the spillway (Figs. 4A and 3D).

We speculate that these fluctuations in outgassing and fountaining at the vent are modulated by a "gas piston" process. Gas pistoning is the cyclic rise and fall of a ponded lava surface, with intense spattering accompanying the fall phase (*32*). The process has been frequently observed at Kīlauea and can be explained as the periodic growth and collapse of foam at the top of the lava column or lava lake (*33*). During peak lava levels of the pulsing regimes, gas release and seismic tremor are inhibited at the vent as gas accumulates in a foamy layer at the top of the lava column, producing weak dome fountaining at the vent and little bubble bursting. The foamy "head" spills out of the crater into the channel, producing a gas-charged, higher-volume flow. Eventually, the foam at the top of the lava column breaks down, liberating accumulated gas and driving a more vigorous fountain at the vent with extensive bubble bursting and higher seismic tremor (RSAM). The lava pouring out of the crater is more efficiently outgassed and has a lower bulk volume, producing low levels of lava in the channel. We can explain the two regimes that fissure 8 underwent as a transition from periods of steady outgassing (nonpulsing regime) to oscillatory gas pistoning (pulsing regime). Similar sporadic regimes of gas pistoning were commonly observed at Kīlauea's summit lava lake during 2008 to 2018 (*34*).

Gas pistoning is only one potential outgassing-driven model to explain the pulsing. A challenge to this model is that gas pistoning is normally observed within a confined pond (*32*, *33*), whereas the fissure 8 pond was feeding a substantial outflow. Also, foam buildup may be difficult to reconcile with the observed low fountaining and surface disruption that occurred in those phases of the cycles (Fig. 4A). Regardless of the model, the data suggest that some type of shallow outgassing process local to the fissure 8 vent modulated the pulsing effusion.

Comparable short-term cycles in bulk effusion rate were observed during the 1984 Mauna Loa eruption and were attributed to a similar gas-driven process. Study of the 1984 fountains and proximal channel showed that during high lava fountaining, the lava was outgassed more effectively, supplying lower-volume, denser lava to the channel and lowering the lava level without changing the lava mass flux (*17*).

On the basis of that previous study (*17*), we speculate that the pulses at fissure 8 did not involve a major change in lava supply rate [i.e., bubble-free "dense rock equivalent" (DRE) effusion rate]. We can investigate this idea by

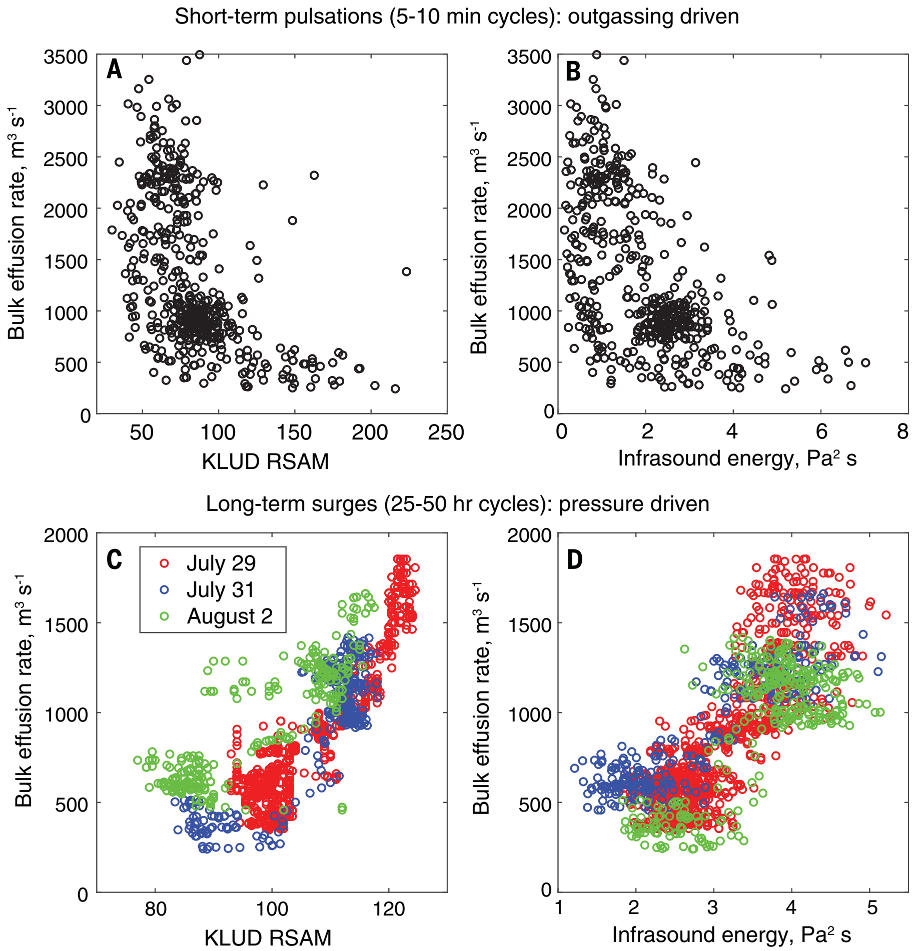


Fig. 5. Relationship of RSAM and bulk effusion rates for two types of cycles. (**A**) Bulk effusion rate showing an inverse correlation with RSAM during short-term pulsations. (**B**) Relationship between bulk effusion rate and infrasound energy during short-term pulsing. (**C**) Bulk effusion rate showing a linear correspondence with RSAM during the longer-term surges. (**D**) Bulk effusion rate showing a correlation with infrasound energy during the longer-term surges.

converting the bulk effusion rates to DRE effusion rates using the vesicularities that we observed from lava samples collected from the spillway (figs. S10 and S12) (30). The measured vesicularity range of 50% (gas-poor lava during pulsing troughs) to 82% (gas-rich lava during pulsing peaks) reduces the wide span of observed bulk effusion rates (350 to 1700 m^3 s^{-1}) (Fig. 2F) during pulsing to a narrower DRE effusion range (175 to 306 m^3 s^{-1}). Although the average DRE values of peak and trough stages remain offset, the uncertainty in the bulk effusion rate values (~15%) (30) (Fig. 2F), combined with the likelihood that the lava samples may not capture the whole range of vesicularity of lava flowing in the channel, precludes confidence of a lava supply rate change. The exercise demonstrates that much of the bulk effusion rate difference could be accounted for by vesicularity changes from the variable degassing that we observed in the UAS video.

Origin of long-term fluctuations (surges)

The short delay (minutes) between the summit collapse events and the onset of increased effusion rates at fissure 8 (Figs. 7 and 8 and figs. S13 and S14) and the 40-km span from the summit to the LERZ eruption site (Fig. 1A) informed us about the process behind the surges. We inferred that the surges were driven by a pressure pulse transmitted down the ERZ conduit, not by the transport of a batch of new magma. The pressure transient would move along the 40-km-long conduit at seismic velocities, whereas migration of the magma itself would require much longer time scales (hours or longer) (35). Although bubbles in the magma would reduce the seismic velocity compared with dense rock, the travel time in this pressure-driven scenario would nevertheless be tens of seconds (36–38).

Although effusion rates began to increase within minutes of summit collapses, surge-driven effusion rates took ~2 to 4 hours to peak (Figs. 7F and 8, D, H, and L). This delay is an important constraint on the process driving surges and might be explained by pressure buffering by an intermediate magma storage zone or zones along the ERZ conduit (39). Pressure buffering in a shallow reservoir was used to explain pressure transients that peaked at Puʻu ʻŌʻō several hours after their onset at the summit, 20 km away, during the recent years of the Puʻu ʻŌʻō eruption (39).

Unlike the short-term pulses, our geophysical and downflow observations offer evidence that the increase in bulk effusion rate associated with surges was reflective of a major increase in lava supply rate. The direct scaling among seismic tremor, infrasound, and bulk effusion rate (Fig. 5, C and D) suggested that an increase in fountain vigor at the vent (the likely driver of much of the tremor and infrasound) accompanied the increase in bulk effusion rate. This relationship was opposite to that of the pulses (Fig. 5, A and B), demonstrating a concurrent increase in fountain and flow activity. Downflow, the increased bulk effusion rate associated with the surges, produced overflows in the distal portion of the flow (movie S6), suggesting a large increase in DRE effusion rate that was sufficient to cause changes kilometers down the channel. The pulses, on the other hand, had no medial or distal effects, consistent with any change in bubble content in the near-vent channel being lost to outgassing rapidly with distance (17, 40).

If we assume that the gas content of lava did not change greatly during the surges, then we can use a constant vesicularity to convert the observed bulk effusion rates to DREs. The main profile of samples from the spillway channel walls has a mean vesicularity of 72% (SD 4%) and we assumed this value. Applying this value to the bulk effusion rates during periods before surges (mean: 548 ± 126 m^3 s^{-1}), we estimated DRE effusion rates of 153 (±35) m^3 s^{-1} for typical rates before the surges. Peak surge levels of bulk effusion rate were 1400 to 1700 m^3 s^{-1}, or ~400 to 500 m^3 s^{-1} for DRE effusion rate. These DRE values during the surge peaks are ~100 times greater than recent effusion rates at Puʻu ʻŌʻō (24, 25) but similar to common Mauna Loa effusion rates (17).

Gas- and pressure-driven fluctuations

The pressure- and gas-driven dichotomy for the fissure 8 channel bears similarity to the behavior of the lava lake in Halemaʻumaʻu during Kīlauea's 2008–2018 summit eruption (34). In the Halemaʻumaʻu lava lake, short-term (minutes to hours) fluctuations in the height of the lava surface were driven by shallow outgassing, specifically gas pistoning (34, 41). These short-term level variations had an inverse relationship with outgassing rates, RSAM, and infrasound, like the fissure 8 vent (Fig. 3C). Longer-term variations in the Halemaʻumaʻu

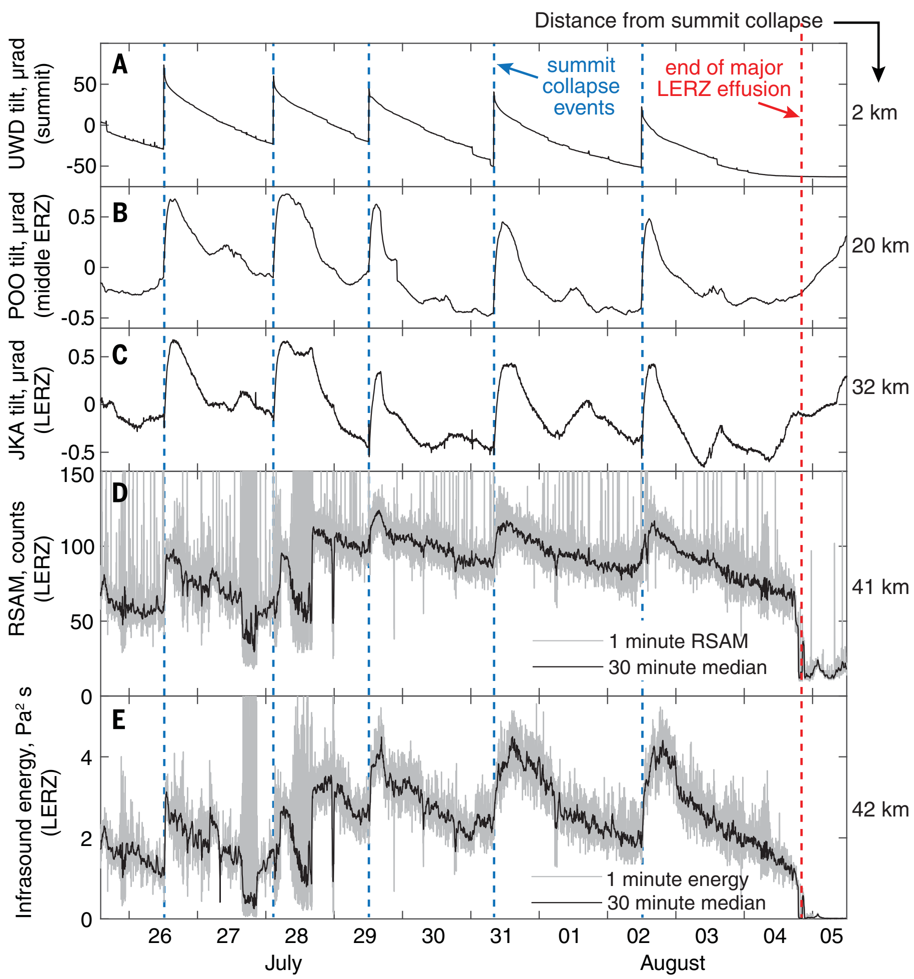


Fig. 6. Ten days of tilt, RSAM, and infrasound spanning five summit collapse events. Summit collapses were registered sharply at summit tilt stations [(**A**), UWD station]. Tilt offsets were also registered tens of kilometers away in the middle ERZ [(**B**), POO station] and LERZ [(**C**), JKA station]. During this period, the summit collapse events were followed by an increase in tremor (shown by peaks in RSAM) in the LERZ [(**D**), from station KLUD; Fig. 1B], suggesting an increase in LERZ eruption vigor. (**E**) Infrasound energy showing peaks after the summit collapses, another indicator of eruption escalation on the LERZ.

lava lake level, lasting hours to days, were driven by magma reservoir pressure (*34*, *42*), again, like the fissure 8 vent.

This shared pattern of pressure- and gas-driven behavior at Halemaʻumaʻu and fissure 8 is noteworthy given the very different eruption styles and might suggest a general starting point for interpreting and forecasting fluctuations at open-vent basaltic volcanoes—short-term (minutes to hours) variations in eruptive behavior are likely to be driven by shallow outgassing fluctuations, whereas longer-term (hours to days) changes are likely related to lava supply rates and reservoir pressure.

Insights into caldera collapse and the magmatic system

Eruption rate variations on Kīlauea's LERZ in 2018 provided us with broader insights into the volcano's magmatic system and, more generally, into rift-summit interactions during basaltic caldera-forming eruptions. Summit collapse–driven surges in lava supply to the LERZ vent bear relevance to two concepts on caldera collapse that have developed in recent decades. First, studies show that there can be a complex interplay between flank eruptions and their parental summit magma reservoirs (*27*, *43*, *44*). Summit reservoirs supply pressurized magma to the flank vents, but the flank conduits regulate the rate of summit reservoir draining, producing a two-way interaction. Second, most of the caldera collapses monitored with modern instrumentation, including those at Miyakejima (Japan), Fernandina (Galapagos), Piton de la Fournaise (Reunion Island), Bárðarbunga (Iceland), and now Kīlauea, have exhibited episodic, and quasiperiodic, progression (*8*, *43*, *45*). Our results confirmed and extend these two concepts to show that the episodic rhythm of summit caldera collapse sequences may be imparted on the accompanying flank eruption, and this episodic flank effusion can have direct implications for hazard.

The postcollapse surges at Kīlauea offered clear evidence of an efficient hydraulic connection between the summit and ERZ (Fig. 1A). A sustained magmatic link between the summit and ERZ was illustrated during the 35-year-long Puʻu ʻŌʻō eruption, and repeated observations showed that transient pressure increases in the summit magma reservoir produced higher eruption rates at the Puʻu ʻŌʻō vents 20 km from the summit (*42*, *46*). The postcollapse surges in 2018 demonstrated that transmission of pressure changes by a subhorizontal magmatic conduit can be sustained over twice that distance, at >40 km. Although previous work has shown that lateral magma transfer can occur at distances of 40+ km (*43*), our results build upon this to show that pressure communication between the summit and flank vent over these distances can occur over time scales as short as minutes. Quantitative estimates of time-variable eruption rates during fissure 8 surges will be a vital component in future efforts to model the magma flow from the summit to the vent and could be used to constrain the properties of the ERZ conduit and the summit reservoir during the collapse events (*47*, *48*).

Tremor as a tool for monitoring eruption rates

Seismic tremor has been correlated with eruption intensity at several volcanoes and across different eruption styles (*49*–*52*). Reliably linking tremor amplitude with the volumetric eruption rate is a valuable objective for operational monitoring and hazard assessment because tracking tremor is much easier than estimating real-time effusion rates. However, robust comparison datasets are rare. Complicating the relationship is that tremor associated with basaltic volcanism is often strongly tied to near-surface outgassing (*41*, *53*, *54*), which may not be indicative of deeper magma supply rates.

If we disregard short-term gas-driven changes (pulses) in RSAM and focus on long-term surges, we find a correlation (R = 0.80) between RSAM and effusion rate (Fig. 5C). This trend suggests that, in some circumstances, tremor can be used to gauge lava supply rates during basaltic eruptions. We note the caveat that outgassing-driven fluctuations must be identified and disregarded.

Infrasonic tremor has also been explored as a tool for monitoring volumetric eruption rates, primarily at explosive volcanoes (*52*, *55*, *56*). For the Kīlauea eruption, both infrasonic and seismic tremor had a direct scaling with lava

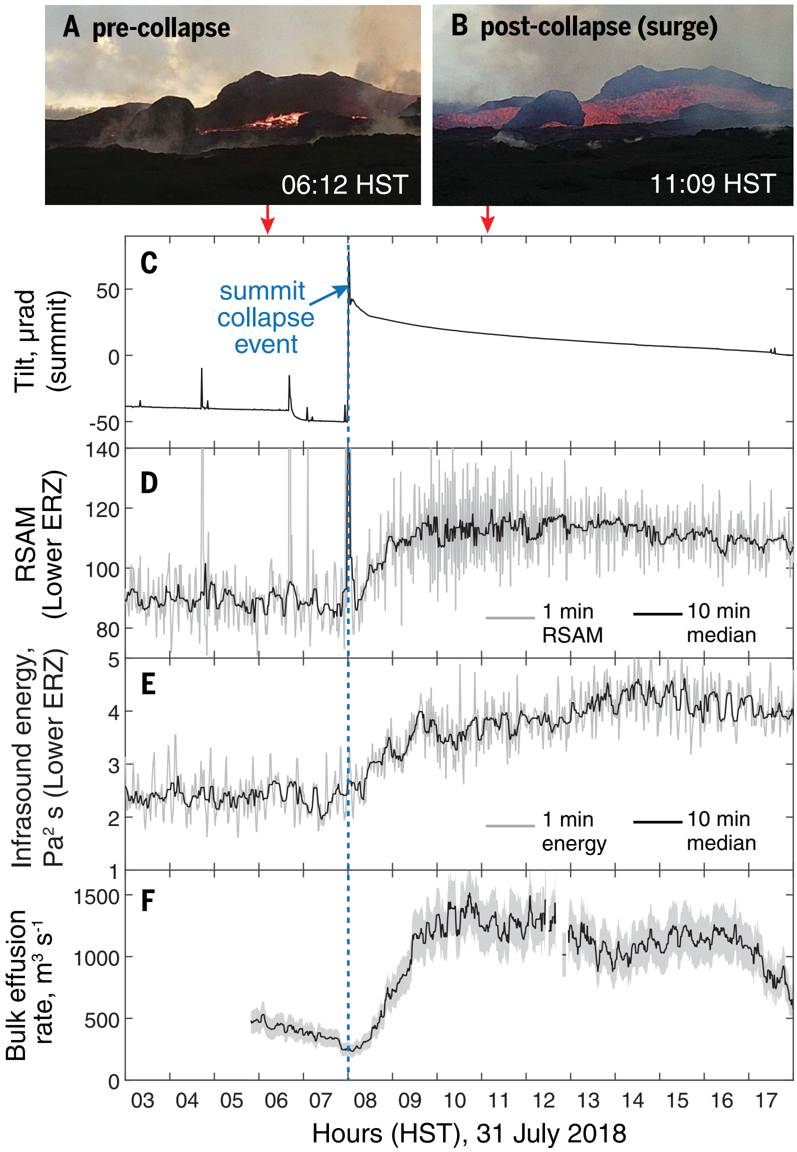


Fig. 7. Example of a LERZ surge event after summit collapse on 31 July 2018. (**A** and **B**) Images of the lava channel before and after the summit collapse event showing a major increase in flow vigor after the event. (**C**) The summit collapse occurred at 08:00 HST, as shown by the tilt offset at the summit. (**D**) RSAM increased within minutes of the summit collapse event, peaking 3to 4 hours after the event. (**E**) Infrasound energy followed the trend in seismic tremor (RSAM). (**F**) Estimated bulk effusion rate from the lava-level data showing an increase of a factor of 2 to 3 after the summit collapse event. Gray area shows the uncertainty in effusion rate estimates based on ±1 m uncertainty in lava level in the channel.

supply rate (Fig. 5D) after excluding the gas-driven pulses (Fig. 5B). Both were largely recording activity at the vent that appears to scale with effusion rate. This demonstrates that infrasonic tremor may be a reliable tool for tracking eruption rates in nonexplosive eruptions.

Hazard implications

The short duration of the pulses in the 2018 LERZ eruption of Kīlauea resulted in their hazard being limited to the proximal region of the flow. Occasional overflows were a threat to evacuated homes adjacent to the flow margin along the first kilometer of the lava channel, and pulsing regimes warranted greater caution in the proximal flow region. Surges, however, produced hazards with a farther reach, as shown on August 2. Then, the increased effusion rate caused lava to rise in the distal channel and overflow the levees, triggering new lobes extending out from the existing flow margins and creating hazards for nearby residents on Noni Farms Road and Papaya Farms Road (Fig. 1B). The observation that the summit collapse events preceded peaks in effusion rate at fissure 8 allowed Hawaiian Volcano Observatory geologists to anticipate potentially hazardous conditions and warn Hawai'i County Civil Defense.

The link between effusion rate and hazard of an advancing flow is well established, as effusion rate is a major control on the flow length and advance rate (*13*). The fissure 8 flow quickly advanced to the ocean and established a relatively stable channel that persisted for 2 months. Although the distal portions of the flow, nearest the coast, changed frequently, the remainder of the channel system sustained high effusion rates with only occasional disruptions. Disruptions, such as channel overflows, coincided with the effusion rate fluctuations driven by the surges. Thus, whereas the absolute effusion rate was important for gauging hazards during initial flow advance, once a persistent channel was established, the lateral hazards were controlled by the variations in effusion rate. Overflows trigger new lobes, generating hazards along the flow margins; they may also trigger levee breaches that reduce supply at the flow front and lessen hazard there (*17*).

Conclusion

The 2018 LERZ eruption of Kīlauea presented an excellent opportunity to study the dynamics of high-effusion lava flows using modern tools. The sustained nature of the fissure 8 flow allowed us to collect a robust, multidisciplinary dataset to examine the diverse processes that drive fluctuations in flow vigor. The two time scales of effusion rate fluctuations corresponded to a shallow, near-vent outgassing process and a deeper, pressure-driven change originating from the episodic caldera-collapse events at the summit, 40 km distant. The hydraulic connection between the summit magma reservoir and the flank eruption allowed the episodic nature of summit collapses to be rapidly expressed as changes in eruption vigor on the flank. The integrated dataset, coupled with frequent direct observations, was essential for understanding the nature and hazard implications of these variations.

Materials and methods

U.S. Geological Survey (USGS) field crews were on the ground 24/7 during the eruption and made frequent direct observations of the proximal fissure 8 channel. Lava flow effusion rates were estimated using constraints on the velocity and cross-sectional area of lava flowing through the proximal channel from ground-based video and time-lapse images. We followed the technique described previously (*13*) to correct for the depth-averaged flow velocity. Bulk effusion rates were converted to DRE (bubble-free) values using the density of lava

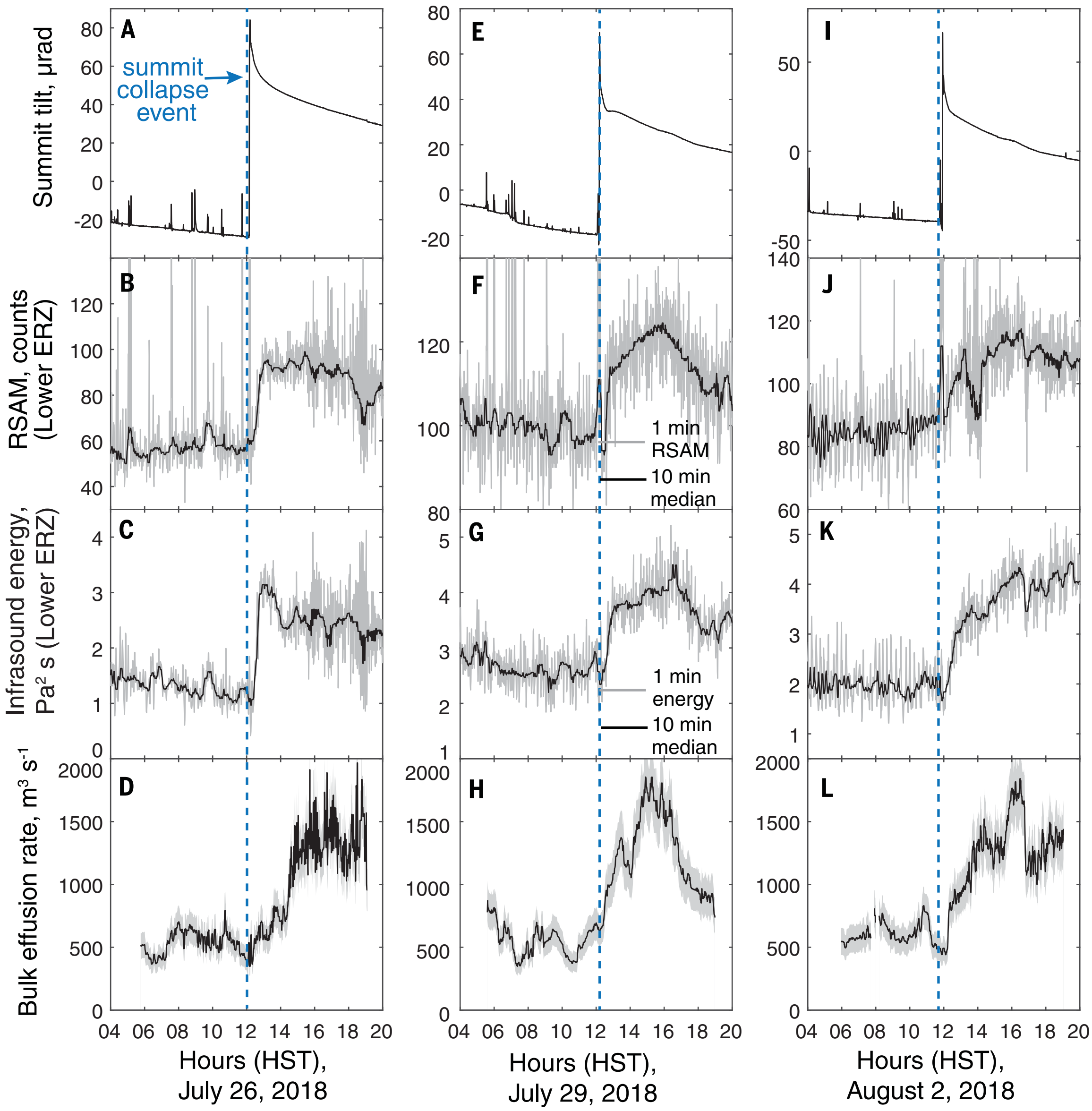


Fig. 8. Other examples of LERZ surges after summit collapse events. (**A**, **E**, and **I**) Summit ground tilt (station UWD) showing the time of summit collapse events. (**B**, **F**, and **J**) LERZ RSAM showing the increase in RSAM after the summit collapse events. (**C**, **G**, and **K**) Infrasound energy at the LERZ vent increased within minutes of the summit collapse events. (**D**, **H**, and **L**) Estimated bulk effusion rates began rising within 20 min of the summit collapse events (*30*) and peaked several hours afterward. Gray area shows the uncertainty in effusion rate estimates based on ±1 m uncertainty in lava level in the channel.

samples collected from the channel walls after the eruption ceased. We also measured variations in channel velocity and vent activity using nadir-viewing UAS video. Thermal images were collected in short campaigns using a handheld thermal camera positioned 300 m from the vent. Seismic tremor was tracked by a permanent seismometer 1 km from the vent, and infrasound was measured by a temporary four-microphone array 500 m from the vent. Summit collapses and the ensuing deformation changes on the ERZ were tracked with several electronic tiltmeters.

REFERENCES AND NOTES

1. A. J. L. Harris, "Basaltic lava flow hazard," in *Volcanic Hazards, Risks, and Disasters*, J. F. Shroder, P. Papale, Eds. (Elsevier, 2015), pp. 17–46.
2. A. Harris, T. De Groeve, S. Carn, F. Garel, Risk evaluation, detection and simulation during effusive eruption disasters. *Geol. Soc. Lond. Spec. Publ.* **426**, 1–22 (2016). doi: 10.1144/SP426.29
3. S. F. Jenkins, S. J. Day, B. V. E. Faria, J. F. B. D. Fonseca, Damage from lava flows: Insights from the 2014–2015 eruption of Fogo, Cape Verde. *J. Appl. Volcanol.* **6**, 6 (2017). doi: 10.1186/s13617-017-0057-6
4. D. Tedesco *et al.*, January 2002 volcano-tectonic eruption of Nyiragongo volcano, Democratic Republic of Congo. *J. Geophys. Res. Solid Earth* **112**, B09202 (2007). doi: 10.1029/2006JB004762
5. L. Stieltjes, P. Moutou, A statistical and probabilistic study of the historic activity of Piton de la Fournaise, Reunion Island, Indian Ocean. *J. Volcanol. Geotherm. Res.* **36**, 67–86 (1989). doi: 10.1016/0377-0273(89)90006-1
6. S. Calvari, M. Coltelli, M. Neri, M. Pompilio, The 1991-1993 Etna eruption chronology and lava flow - field evolution. *Acta Vulcanol.* **4**, 1–14 (1994).
7. D. Andronico *et al.*, A multi-disciplinary study of the 2002-03 Etna eruption: Insights into a complex plumbing system. *Bull. Volcanol.* **67**, 314–330 (2005). doi: 10.1007/s00445-004-0372-8
8. C. A. Neal *et al.*, The 2018 rift eruption and summit collapse of Kīlauea Volcano. *Science* **363**, 367–374 (2019). pmid: 30538164
9. M. Poland *et al.*, The 2014–2015 Pāhoa lava flow crisis at Kilauea Volcano, Hawai'i: Disaster avoided and lessons learned. *GSA Today* **26**, 4–10 (2016). doi: 10.1130/GSATG262A.1
10. C. Heliker, T. N. Mattox, "The first two decades of the Pu'u 'O'o-Kupaianaha eruption: Chronology and selected

bibliography" in *The Pu'u 'O'o-Kupaianaha Eruption of Kilauea Volcano, Hawaii: The First 20 Years*, C. C. Heliker, D. A. Swanson, T. J. Takahashi, Eds. (Professional Paper 1676, U.S. Geological Survey, 2003), pp. 1–27.
11. A. Vicari *et al*., Near-real-time forecasting of lava flow hazards during the 12-13 January 2011 Etna eruption. *Geophys. Res. Lett.* **38**, 1–7 (2011). doi: 10.1029/2011GL047545
12. G. P. L. Walker, A. T. Huntingdon, A. T. Sanders, J. L. Dinsdale, Lengths of lava flows. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **274**, 107–118 (1973). doi: 10.1098/rsta.1973.0030
13. A. J. L. Harris, J. Dehn, S. Calvari, Lava effusion rate definition and measurement: A review. *Bull. Volcanol.* **70**, 1–22 (2007). doi: 10.1007/s00445-007-0120-y
14. A. J. L. Harris, S. K. Rowland, Flowgo 2012: An updated framework for thermorheological simulations of channel-contained lava. *Geophys. Monogr. Ser.* **208**, 457–481 (2015). doi: 10.1002/9781118872079.ch21
15. C. Del Negro, L. Fortuna, A. Herault, A. Vicari, Simulations of the 2004 lava flow at Etna volcano using the magflow cellular automata model. *Bull. Volcanol.* **70**, 805–812 (2008). doi: 10.1007/s00445-007-0168-8
16. S. Tarquini, M. de' Michieli Vitturi, Influence of fluctuating supply on the emplacement dynamics of channelized lava flows. *Bull. Volcanol.* **76**, 801 (2014). doi: 10.1007/s00445-014-0801-2
17. P. W. Lipman, N. Banks, "A'a flow dynamics, 1984 Mauna Loa eruption," in *Eruptions of Hawaiian Volcanoes: Past, Present, and Future*, R. I. Tilling, C. Heliker, D. A. Swanson, Eds. (Professional Paper 1350, U. S. Geological Survey, 1987), pp. 1527–1567.
18. J. E. Bailey, A. J. L. Harris, J. Dehn, S. Calvari, S. K. Rowland, The changing morphology of an open lava channel on Mt. Etna. *Bull. Volcanol.* **68**, 497–515 (2006). doi: 10.1007/s00445-005-0025-6
19. M. R. James, L. J. Applegarth, H. Pinkerton, Lava channel roofing, overflows, breaches and switching: Insights from the 2008-2009 eruption of Mt. Etna. *Bull. Volcanol.* **74**, 107–117 (2012). doi: 10.1007/s00445-011-0513-9
20. R. I. Tilling, J. P. Kauahikaua, S. R. Brantley, C. A. Neal, "The Hawaiian Volcano Observatory: A natural laboratory for studying basaltic volcanism," in *Characteristics of Hawaiian Volcanoes*, M. P. Poland, T. J. Takahashi, C. M. Landowski, Eds. (Professional Paper 1801-1, U.S. Geological Survey, 2014), pp. 1–64.
21. M. Zoeller, M. Patrick, C. Neal, SectorInsight.gov–Crisis Remote Sensing during the 2018 Lower East Rift Zone Eruption of K lauea Volcano. *Photogramm. Eng. Remote Sensing* **84**, 749–751 (2018). doi: 10.14358/PERS.84.12.749
22. T. L. Wright, F. W. Klein, *Two Hundred Years of Magma Transport and Storage at Kīlauea Volcano, Hawai'i, 1790–2008* (Professional Paper 1806, U.S. Geological Survey, 2014).
23. T. R. Orr *et al*., Kilauea's 5–9 March 2011 Kamoamoa fissure eruption and its relation to 30+ years of activity from Pu'u 'Ō 'ō. *Geophys. Monogr. Series* **208**, 393–420 (2015). doi: 10.1002/9781118872079.ch18
24. J. Kauahikaua, D. R. Sherrod, K. V. Cashman, C. Heliker, K. Hon, T. N. Mattox, J. A. Johnson, "Hawaiian lava-flow dynamics during the Pu'u 'O'o-Kupaianaha eruption: A tale of two decades," in *The Pu'u 'O'o-Kupaianaha Eruption of Kilauea Volcano, Hawaii: The First 20 Years*, C. C. Heliker, D. A. Swanson, T. J. Takahashi, Eds. (Professional Paper 1676, U.S. Geological Survey, 2003), pp. 63–88.
25. M. P. Poland, Time-averaged discharge rate of subaerial lava at Kīlauea Volcano, Hawai'i, measured from TanDEM-X interferometry: Implications for magma supply and storage during 2011-2013. *J. Geophys. Res. Solid Earth* **119**, 5464–5481 (2014). doi: 10.1002/2014JB011132
26. D. Wilson *et al*., Small explosion from new vent at Kilauea's summit. *Eos* **89**, 203 (2008). doi: 10.1029/2008EO220003
27. M. Patrick, T. Orr, K. Anderson, D. Swanson, Eruptions in sync: Improved constraints on Kīlauea Volcano's hydraulic connection. *Earth Planet. Sci. Lett.* **507**, 50–61 (2019). doi: 10.1016/j.epsl.2018.11.030
28. H. R. Dietterich, M. R. Patrick, A. K. Diefenbach, C. Parcheta, E. Lev, N. L. Foks, "Lava flow hazard modeling and the assessment of effusion rates and topographic change with UAS and lidar during the 2018 Kilauea lower East Rift Zone eruption," Paper presented at the 2018 American Geophysical Union Fall Meeting, Washington, D.C. (2018), abstract #V21B-03.
29. E. T. Endo, T. L. Murray, Real-time seismic amplitude measurement (RSAM): A volcano monitoring and prediction tool. *Bull. Volcanol.* **53**, 533–545 (1991). doi: 10.1007/BF00298154
30. Materials and methods are available as supplementary materials.
31. M. R. Patrick, T. R. Orr, Rootless shield and perched lava pond collapses at Kīlauea Volcano, Hawai'i. *Bull. Volcanol.* **74**, 67–78 (2011). doi: 10.1007/s00445-011-0505-9
32. D. A. Swanson, W. A. Duffield, D. B. Jackson, D. W. Peterson, *Chronological Narrative of the 1969–71 Mauna Ulu Eruption of Kilauea Volcano, Hawaii* (Professional Paper 1056, U.S. Geological Survey, 1979).
33. T. R. Orr, J. C. Rea, Time-lapse camera observations of gas piston activity at Pu'u 'O'o, Kilauea volcano, Hawai'i. *Bull. Volcanol.* **74**, 2353–2362 (2012). doi: 10.1007/s00445-012-0667-0
34. M. Patrick, D. Swanson, T. Orr, A review of controls on lava lake level: Insights from Halema'uma'u Crater, Kīlauea Volcano. *Bull. Volcanol.* **81**, 13 (2019). doi: 10.1007/s00445-019-1268-y
35. M. Poland, A. Miklius, E. K. Montgomery-Brown, "Magma supply, storage, and transport at shield-stage Hawaiian volcanoes," in *Characteristics of Hawaiian Volcanoes*, M. P. Poland, T. J. Takahashi, C. M. Landowski, Eds. (Professional Paper 1801-1, U.S. Geological Survey, 2014), pp. 149–234.
36. H. Kumagai, B. A. Chouet, Acoustic properties of a crack containing magmatic or hydrothermal fluids. *J. Geophys. Res. Solid Earth* **105** (B11), 25493–25512 (2000). doi: 10.1029/2000JB900273
37. S. Sturton, J. Neuberg, The effects of a decompression on seismic parameter profiles in a gas-charged magma. *J. Volcanol. Geotherm. Res.* **128**, 187–199 (2003). doi: 10.1016/S0377-0273(03)00254-3
38. L. Collier, J. W. Neuberg, N. Lensky, V. Lyakhovsky, O. Navon, Attenuation in gas-charged magma. *J. Volcanol. Geotherm. Res.* **153**, 21–36 (2006). doi: 10.1016/j.jvolgeores.2005.08.009
39. M. Patrick, M. Haney, K. R. Anderson, "Ground tilt time delays between Kilauea Volcano's summit and east rift zone caused by magma reservoir buffering," Paper presented at the 2016 American Geophysical Union Fall Meeting, San Francisco (2016), abstract #V12A-06.
40. J. M. Riker, K. V. Cashman, J. P. Kauahikaua, C. M. Montierth, The length of channelized lava flows: Insight from the 1859 eruption of Mauna Loa Volcano, Hawai'i. *J. Volcanol. Geotherm. Res.* **183**, 139–156 (2009). doi: 10.1016/j.jvolgeores.2009.03.002
41. M. R. Patrick *et al*., Shallowly driven fluctuations in lava lake outgassing (gas pistoning), Kilauea Volcano. *Earth Planet. Sci. Lett.* **433**, 326–338 (2016). doi: 10.1016/j.epsl.2015.10.052
42. M. R. Patrick, K. R. Anderson, M. P. Poland, T. R. Orr, D. A. Swanson, Lava lake level as a gauge of magma reservoir pressure and eruptive hazard. *Geology* **43**, 831–834 (2015). doi: 10.1130/G36896.1
43. M. T. Gudmundsson *et al*., Gradual caldera collapse at Bárdarbunga volcano, Iceland, regulated by lateral magma outflow. *Science* **353**, aaf8988 (2016). doi: 10.1126/science.aaf8988; pmid: 27418515
44. M. Ripepe *et al*., Forecasting effusive dynamics and decompression rates by magmastatic model at open-vent volcanoes. *Sci. Rep.* **7**, 3885 (2017). doi: 10.1038/s41598-017-03833-3; pmid: 28634369
45. L. Michon, N. Villeneuve, T. Catry, O. Merle, How summit calderas collapse on basaltic volcanoes: New insights from the April 2007 caldera collapse of Piton de la Fournaise volcano. *J. Volcanol. Geotherm. Res.* **184**, 138–151 (2009). doi: 10.1016/j.jvolgeores.2008.11.003
46. T. R. Orr, J. E. Bleacher, M. R. Patrick, K. M. Wooten, A sinuous tumulus over an active lava tube at Kīlauea Volcano: Evolution, analogs, and hazard forecasts. *J. Volcanol. Geotherm. Res.* **291**, 35–48 (2015). doi: 10.1016/j.jvolgeores.2014.12.002
47. K. Anderson, M. Poland, J. H. Johnson, A. Miklius, Episodic deflation-inflation events at Kilauea Volcano and implications for the shallow magma system. *Hawaiian volcanoes, from source to surface*. *Geophys. Monogr.* **208**, 229–250 (2015). doi: 10.1002/9781118872079.ch11
48. K. R. Anderson, M. P. Poland, Bayesian estimation of magma supply, storage, and eruption rates using a multiphysical volcano model: Kīlauea Volcano, 2000–2012. *Earth Planet. Sci. Lett.* **447**, 161–171 (2015). doi: 10.1016/j.epsl.2016.04.029
49. G. A. Macdonald, J. P. Eaton, Hawaiian volcanoes during 1955. *U.S. Geol. Surv. Bull.* **1171**, 167 (1964).
50. J. Battaglia, K. Aki, T. Staudacher, Location of tremor sources and estimation of lava output using tremor source amplitude on the Piton de la Fournaise volcano: 2. Estimation of lava output. *J. Volcanol. Geotherm. Res.* **147**, 291–308 (2005). doi: 10.1016/j.jvolgeores.2005.04.006
51. D. Bertin *et al*., High effusion rates of the Cordõn Caulle 2011-2012 eruption (Southern Andes) and their relation with the quasi-harmonic tremor. *Geophys. Res. Lett.* **42**, 7054–7063 (2015). doi: 10.1002/2015GL064624
52. D. Fee *et al*., Volcanic tremor and plume height hysteresis from Pavlof Volcano, Alaska. *Science* **355**, 45–48 (2017). doi: 10.1126/science.aah6108; pmid: 28059760
53. J. L. Palma, E. S. Calder, D. Basualto, S. Blake, D. Rothery, Correlations between SO_2 flux, seismicity, and outgassing activity at the open vent of Villarrica volcano, Chile. *J. Geophys. Res. Solid Earth* **113** (B10), B10201 (2008). doi: 10.1029/2008JB005577
54. G. G. Salerno, M. Burton, G. Di Grazia, T. Caltabiano, C. Oppenheimer, Coupling between magmatic degassing and volcanic tremor in basaltic volcanism. *Front. Earth Sci.* **6**, 157 (2018). doi: 10.3389/feart.2018.00157
55. D. Fee *et al*., Eruption mass estimation using infrasound waveform inversion and ash and gas measurements: Evaluation at Sakurajima Volcano, Japan. *Earth Planet. Sci. Lett.* **480**, 42–52 (2017). doi: 10.1016/j.epsl.2017.09.043
56. S. De Angelis, A. Diaz-Moreno, L. Zuccarello, Recent developments and applications of acoustic infrasound to monitor volcanic emissions. *Remote Sens.* **11**, 1301 (2019). doi: 10.3390/rs11111302
57. M. R. Patrick, H. R. Dietterich, J. J. Lyons, A. K. Diefenbach, C. Parcheta, K. R. Anderson, A. Namiki, I. Sumita, B. Shiro, J. P. Kauahikaua, "Cyclic lava effusion during the 2018 eruption of Kīlauea Volcano: data release," in *U.S. Geological Survey Data Release* (2019); https://doi.org/10.5066/P9PJZ17R.

ACKNOWLEDGMENTS

We thank Hawai'i County Civil Defense for facilitating USGS work within restricted zones of the eruption site; USGS staff from outside HVO for assistance in making ground observations; Dr. David Fee for the infrasound stations used in this study; Sanford's Service Center for permission to deploy a time-lapse camera at their Kapoho quarry; the University of Hawai'i at Hilo geology department for hosting USGS scientists after the evacuation of the HVO building; and Leilani Estates residents for support during the eruption crisis. **Funding:** The 2018 eruption response was funded by the USGS Volcano Science Center. A.N. was supported by KAKENHI 17KK0092 and I.S. was supported by Kanazawa University and KAKENHI 15K13591. **Author contributions:** M.P. wrote the manuscript and collected and interpreted the field camera data; H.D. contributed to effusion rate estimates, interpretation, and writing; J.L. processed and interpreted the infrasound; A.D. oversaw collection of the UAS data; C.P., K.A., A.N., I.S., B.S., and J.K. made field observations and contributed to interpretations. **Competing interests:** The authors declare no competing interests. The use of brand names is for information only and does not imply endorsement by the Federal Government. **Data and materials availability:** The data used in this study will be available upon publication in the USGS ScienceBase online catalog (*57*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaay9070/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
References (*58*–*60*)
Movies S1 to S6

3 August 2019; accepted 11 November 2019
10.1126/science.aay9070

RESEARCH ARTICLE SUMMARY

VOLCANOLOGY

Magma reservoir failure and the onset of caldera collapse at Kīlauea Volcano in 2018

Kyle R. Anderson*, Ingrid A. Johanson, Matthew R. Patrick, Mengyang Gu, Paul Segall, Michael P. Poland, Emily K. Montgomery-Brown, Asta Miklius

INTRODUCTION: The 2018 rift zone eruption of Kīlauea Volcano, Hawai'i, drained large volumes of magma from the volcano's summit reservoir system, causing high-rate subsidence of the ground surface and withdrawal of an active lava lake. Over the span of 1 week, the surface of the lava lake fell more than 300 m. Continued withdrawal of magma caused the rock above the reservoir to fail, triggering the onset of episodic caldera collapse. Surface collapse began near the evacuated lava lake vent, but as the eruption continued over 3 months, the area of the new caldera expanded to ~5 km^2 and its volume grew to 0.8 km^3. The precursory activity and subsequent growth of the caldera were recorded in far greater detail than was possible at the handful of other caldera collapses observed in the past century. These comprehensive observations permit new insights into the conditions that lead to magma reservoir host rock failure and caldera collapse.

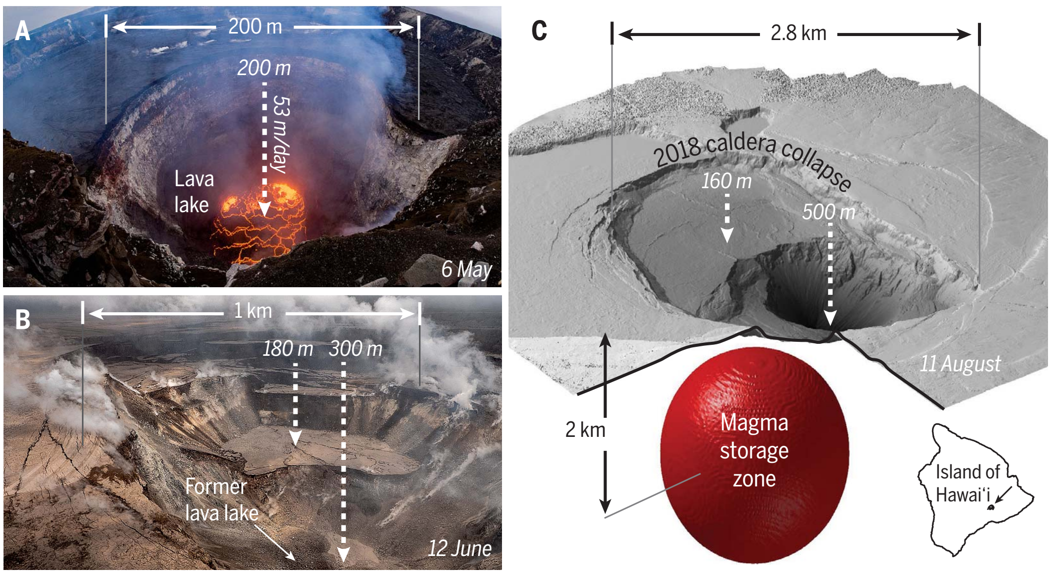
RATIONALE: Volcanic caldera collapses can be highly destructive and create prominent topographic features, but little is known about the architecture of subcaldera magma storage zones or the critical decrease in pressure that triggers collapse. Withdrawal of Kīlauea's lava lake in 2018 can be used to gauge pressure change in the underlying magma reservoir. We developed a model of time-evolving reservoir depressurization to jointly explain lava lake withdrawal rate and the rate and spatial pattern of ground subsidence obtained from radar satellites and a dense local monitoring network.

RESULTS: We tracked the evolution of the magmatic system from steady elastic decompression to inelastic failure. We were able to estimate the location, geometry, volume, and time-evolving pressure within the reservoir as well as conditions required to trigger failure of the overlying crust. Before the onset of collapse, the ground at Kīlauea's summit was subsiding at nearly 10 cm/day, and the lava lake surface was retreating at ~50 m/day. We found that these phenomena were caused by drainage of magma at a high rate from a storage reservoir centered ~2 km below the surface, with a volume of several cubic kilometers. Drainage rapidly reduced reservoir pressure, stressing the surrounding crust. Two weeks after the rift zone intrusion and eruption began to drain magma from the summit, withdrawal of <4% of the stored magma had reduced pressure in the reservoir by ~17 MPa, causing the host rock above it to begin to fail episodically. The episodic collapses loaded the magma with the weight of the roof, increasing its pressure. The final collapse caldera was closely centered over the magma reservoir, and their horizontal dimensions were comparable. However, the estimated reservoir volume was substantially greater than the caldera volume, indicating incomplete evacuation at the end of the eruption.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaz1822

CONCLUSION: Our results tightly constrain the pressure decrease in the magma reservoir before the onset of collapse. Together with geodetic data, this bounds the magma storage volume and the stress changes needed to cause failure of the host rock above the reservoir. Our results demonstrate that a magma reservoir's roof may begin to fail after withdrawal of only a small fraction of the stored magma. At Kīlauea, this process was likely influenced by a relatively thin and wide reservoir roof and preexisting crustal weaknesses, including an established caldera ring-fault system and the lava lake vent. Roof collapses maintained magma pressure, sustaining the eruption, but they did not (as is sometimes assumed) completely repressurize the reservoir. This indicates residual frictional strength on the collapse-bounding faults. The eruption was not terminated by complete evacuation of stored magma, contrary to assumptions sometimes made when interpreting data from past caldera collapses, and indicates that a different process was responsible for the cessation of the eruption. Joint monitoring of ground deformation and lava lake elevation at other volcanoes, when possible, may yield rich insights into magmatic processes and conditions. ■



Caldera collapse at Kīlauea in 2018. (**A**) Precollapse lava lake on 6 May 2018. The lake surface had fallen ~200 m since the onset of the eruption. (**B**) Aerial photograph looking west across Kīlauea's summit on 12 June, after the onset of caldera collapse. Parts of the crater floor had subsided as much as ~180 m as intact blocks. (**C**) Estimated magma storage zone that partially collapsed to form the caldera. Shown is the isosurface enclosing the region that contained magma in our simulations, at 95% confidence. View is to the southeast.

PHOTOS: K. R. ANDERSON, U.S. GEOLOGICAL SURVEY

The list of author affiliations is available in the full article online.
*Corresponding author.
Email: kranderson@usgs.gov
Cite this article as K. R. Anderson *et al.*, *Science* **366**, eaaz1822 (2019).
DOI: 10.1126/science.aaz1822

RESEARCH ARTICLE

VOLCANOLOGY

Magma reservoir failure and the onset of caldera collapse at Kīlauea Volcano in 2018

Kyle R. Anderson[1]*, Ingrid A. Johanson[2], Matthew R. Patrick[2], Mengyang Gu[3], Paul Segall[4], Michael P. Poland[5], Emily K. Montgomery-Brown[1], Asta Miklius[2]

Caldera-forming eruptions are among Earth's most hazardous natural phenomena, yet the architecture of subcaldera magma reservoirs and the conditions that trigger collapse are poorly understood. Observations from the formation of a 0.8–cubic kilometer basaltic caldera at Kīlauea Volcano in 2018 included the draining of an active lava lake, which provided a window into pressure decrease in the reservoir. We show that failure began after <4% of magma was withdrawn from a shallow reservoir beneath the volcano's summit, reducing its internal pressure by ~17 megapascals. Several cubic kilometers of magma were stored in the reservoir, and only a fraction was withdrawn before the end of the eruption. Thus, caldera formation may begin after withdrawal of only small amounts of magma and may end before source reservoirs are completely evacuated.

A volcanic caldera is a topographic depression formed by fault-bounded subsidence or collapse of Earth's surface as magma is withdrawn from a crustal storage reservoir, causing the overlying rock to founder (*1*). Caldera formation can be triggered by magma withdrawal to feed violent explosive eruptions or by intrusion of magma into surrounding rock, sometimes feeding long-lived effusive lava flows. Calderas can be prominent topographic features measuring tens of kilometers in diameter.

Our understanding of volcanic caldera collapses has been strongly limited by a lack of well-documented caldera-forming eruptions. From 1900 to the beginning of 2018, only seven caldera collapses were clearly documented on Earth (*2*, *3*), mostly with limited geophysical and observational networks. Even the well-recorded 2014–2015 collapse at Bárðarbunga, Iceland, occurred beneath hundreds of meters of ice, preventing direct observation (*3*).

The 825 million m^3 caldera collapse at Kīlauea Volcano in 2018 was the largest at the volcano in more than two centuries and was tracked by a dense multiparametric monitoring network and through direct visual observations. These detailed datasets record the transition from steady elastic subsidence to fault-bounded collapse as the roof of Kīlauea's summit reservoir failed in response to high-rate magma withdrawal to supply the volcano's East Rift Zone (ERZ) intrusion and eruption. In this study, we modeled ground deformation and lava lake data to infer properties of the magma system as it evolved toward collapse in May 2018. The data offer direct evidence of pressure change in the magma reservoir and present an opportunity to resolve the volcano's subcaldera magma storage architecture and its relation to collapse timing, style, and volume.

[1]U.S. Geological Survey, California Volcano Observatory, Moffett Field, CA, USA. [2]U.S. Geological Survey, Hawaiian Volcano Observatory, Hilo, HI, USA. [3]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, USA. [4]Department of Geophysics, Stanford University, Stanford, CA, USA. [5]U.S. Geological Survey, Cascades Volcano Observatory Vancouver, WA, USA.
*Corresponding author. Email: kranderson@usgs.gov

Kīlauea Volcano and the 2018 eruption

Kīlauea Volcano, on the island of Hawai'i (Fig. 1), is one of the world's most active volcanoes and erupted almost continuously from 1983 to 2018. For most of that period, Kīlauea's mantle-derived magma supply largely passed through its summit reservoir system before migrating subhorizontally down the volcano's ERZ to erupt as lava flows ~20 km from the summit at or near the Pu'u 'Ō'ō vent.

Beginning in 2008, a lava lake was active at the summit of the volcano within Halema'uma'u crater; by April 2018 its surface area had grown to more than 40,000 m^2. The lava lake was supplied from a shallow magma storage zone (here termed the Halema'uma'u reservoir) hypothesized to exist 1 to 2 km beneath Kīlauea's existing summit caldera (formed in ~1500 CE). Variations in the surface height of the lava lake were strongly correlated with ground deformation, indicating that both were caused by pressure changes in the underlying magma reservoir. Thus, Kīlauea's lava lake acted as a magma reservoir pressure gauge (*4*–*6*).

Kīlauea's 35-year-long eruption ended spectacularly on 30 April 2018 with the intrusion of a dike downrift from Pu'u 'Ō'ō into the volcano's lower ERZ (LERZ) (*7*) (Fig. 1B). On 3 May, the intrusion emerged in the Leilani Estates subdivision, more than 40 km from the volcano's summit, ultimately erupting >1 km^3 of lava and destroying hundreds of homes. The intrusion and eruption triggered wholesale draining of Kīlauea's magma system, from the middle ERZ to the summit. Magma drainage from the summit led to lava lake withdrawal and vent collapse, a series of explosions, and ultimately the formation of a new caldera nested within the larger 1500 CE caldera. Summit collapse and most LERZ lava effusion ended in August 2018 after 3 months.

Magma evacuation and the onset of caldera collapse

We recorded subsidence and later collapse of the ground surface at Kīlauea's summit by visual observations, continuous Global Navigation Satellite System (GPS) stations, borehole tiltmeters, and interferometric synthetic aperture radar (InSAR) interferograms derived from satellite data (*8*) (Figs. 2 to 4). Variation in lava lake surface height was recorded by laser rangefinder, thermal camera imagery, and structure-from-motion photogrammetry (Figs. 3 and 4) (*9*).

Before the onset of the LERZ intrusion, Kīlauea's lava lake had been overflowing onto the floor of Halema'uma'u crater. Deflation began in earnest on 2 May with subsidence and contraction of the ground surface and withdrawal of the lava lake at a rate that reached ~40 m/day (Fig. 3 and fig. S8). On 4 May, an earthquake with moment magnitude (M_w) of 6.9 (M6.9) on the basal decollement between the volcanic pile and the oceanic crust underlying Kīlauea's south flank (*7*, *10*) shook the volcano and produced long-wavelength extensional strain across the summit. By the end of the day, lake withdrawal had accelerated to 53 m/day, and the ground tilt rate at summit instruments had more than doubled (*8*) (Fig. 4). Subsidence continued over the following days in a broad, roughly circular region centered near the east rim of Halema'uma'u at rates of up to nearly 10 cm/day (Fig. 2). Ground deformation and lava lake surface height were highly correlated (Fig. 3D). Before the M_w 6.9 earthquake, we observed ~5 m of lava lake withdrawal for every microradian of caldera-directed ground tilt at station UWE [located near the U.S. Geological Survey (USGS) Hawaiian Volcano Observatory (HVO); Fig. 2], in agreement with observations made over many years at Kīlauea (*4*, *5*). After the earthquake, this ratio had decreased by ~40%.

Rapid withdrawal of the lava lake was accompanied by sporadic explosions as unsupported conduit wall rock fell into the vent (Fig. 1), gradually increasing its diameter (Fig. 4). By 10 May, after dropping more than 300 m in just over a week (supplementary movie S1), the lava lake had disappeared from view and the vent was blocked by rubble. Ground subsidence continued, however,

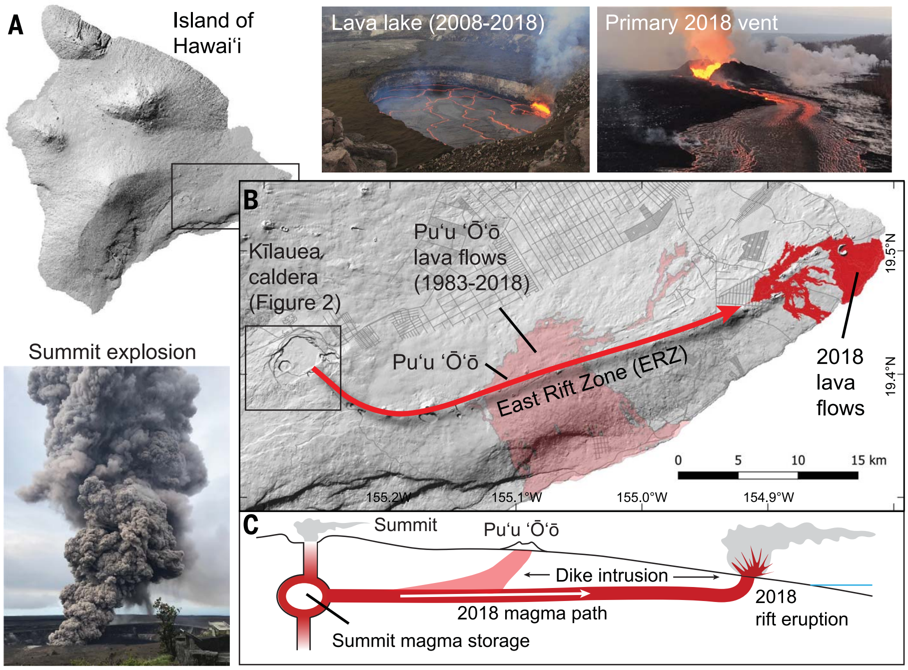


Fig. 1. Kīlauea Volcano and the 2018 eruption. Photos show a summit explosion on 9 May 2018, the lava lake as it appeared in April 2018, and the primary 2018 LERZ eruptive vent. (**A**) Shaded topographic map of the island of Hawai'i; the box shows the extent of the map in (B). (**B**) During the 2018 eruption, magma flowed >40 km underground subhorizontally from the summit (left) to the LERZ vents (right). See Fig. 2 for an enlargement of the summit area. (**C**) Schematic cross section (not to scale) showing flow of magma from the summit to the LERZ.

indicating ongoing depressurization, and HVO became concerned about failure of the rock above the reservoir. From 9 to 15 May, several $M \geq 3$ earthquakes per day shook the summit, and tremor [as indicated by the real-time seismic amplitude measurement (RSAM)] was recorded at very high levels. Ground cracks were observed near Halema'uma'u crater on 14 May, and by 16 May the GPS network had recorded total subsidence in that area of ~1 m.

On 16 May at 18:16 Hawaii Standard Time (HST), abrupt inflationary (radially outward) ground deformation and very-long-period (VLP) seismic energy (M_w 4.9) were recorded across the summit, an ashy gas plume rose to 20,000 ft, and summit RSAM dropped precipitously. Ground deformation and VLP observations were similar to those previously caused by rockfalls into the lava lake and ascribed to pressurization of the shallow magma system (*11*) but were much larger in amplitude. They were also similar to observations recorded during caldera collapses at Miyakejima (Japan) and Piton de la Fournaise (La Réunion) volcanoes (*12*–*15*). Eleven more of these events, informally termed "collapse/explosions" by HVO, occurred before the end of the month. Satellite observations and failure of instruments on the crater rim indicated that the (now empty) lava lake vent was growing more rapidly (Fig. 4) and beginning to cause failure outside of Halema'uma'u, but broader-scale, fault-bounded surface collapse was not yet detected. Summit SO_2 emission rates increased by two to three times (*7*), but erupted tephra volumes were much smaller than collapse volumes. Away from the widening vent, the summit continued to subside between collapses in a roughly circular pattern centered on the caldera.

The onset of broader-scale, clearly fault-bounded collapse outside of Halema'uma'u crater began in the early morning of 29 May with an abrupt down-dropping of the caldera floor around Halema'uma'u, approximately coincident with the onset of higher eruption rates (~150 m^3/s) in the LERZ. We measured 1.5 m of subsidence at a GPS station (NPIT) on the northeast rim of the crater during the seconds-long event, and visual observations from HVO revealed subsidence north-northeast and west of Halema'uma'u. Away from the subsiding block(s), however, inflationary radially outward deformation and VLP seismicity were observed that were similar to previous events in May but with much larger amplitudes (Fig. 4).

On 1 June, enabled by a marked reduction of Kīlauea's summit plume, an unoccupied aerial vehicle took the first clear photos of Halema'uma'u since mid-May. The photos showed major collapse and widening of the vent, ~30 m of subsidence of the western floor of Halema'uma'u, and faulting and subsidence of the 1500 CE caldera floor more than 1 km northwest of the former lava lake. As more collapses occurred through June, the surface expression and area of slumping expanded greatly. Collapse events were roughly periodic in time (Fig. 4A), preceded by marked increases in earthquake rate (*7*), and sometimes followed by surges in effusion rate at the LERZ vent ~40 km distant (*16*). The final collapse geometry was not fully established until mid- to late June, with clockwise propagation of a fault scarp through the center of the older 1500 CE caldera. By the time the new caldera stopped growing in early August, 62 collapses had occurred, producing as much as ~500 m of subsidence and a total collapse area of ~5 km^2.

Modeling lava lake and ground deformation data

Our goals were to estimate the subcaldera magma reservoir geometry; infer the conditions under which the reservoir's host rock began to fail; and evaluate how these parameters related to the style, location, and volume of subsequent caldera collapse. We used data from the period of near-constant high-rate subsidence after the M_w 6.9 earthquake and preceding the first collapse event on 16 May (Fig. 4), which we treated as the effective onset of caldera collapse. Observations suggest that during this time, rock at the summit responded elastically to changing stresses and slip on buried ring faults was minimal (*8*). We hypothesized that ground deformation and changes in lava lake surface height were generated by pressure change at constant rate $\dot{p}$ in a magma reservoir beneath Kīlauea's summit (*4*, *6*). We constructed a model that relates $\dot{p}$ to the rate of lava lake surface height change, assuming a magmastatic relationship, and to observed ground deformation velocities by using a continuum-mechanical model of a spheroidal magma reservoir embedded in an elastic half-space (Fig. 5) (*8*). The deformation model was implemented using the finite element method and then employed to construct a fast numerical surrogate suitable for Markov chain Monte Carlo (MCMC) estimation (*8*, *17*). Primary model parameters are shown in Fig. 5.

We performed a joint Bayesian parameter estimation using the lava lake withdrawal rate together with GPS, ground tilt, and InSAR velocities (*8*). We also used independent information from previous studies to constrain lava lake density and host rock rigidity, and we placed limits on the proximity of the top of the magma reservoir to the surface. We directly estimated reservoir location, geometry, and pressure change rate, and allowed "nuisance" parameters (including host rock shear

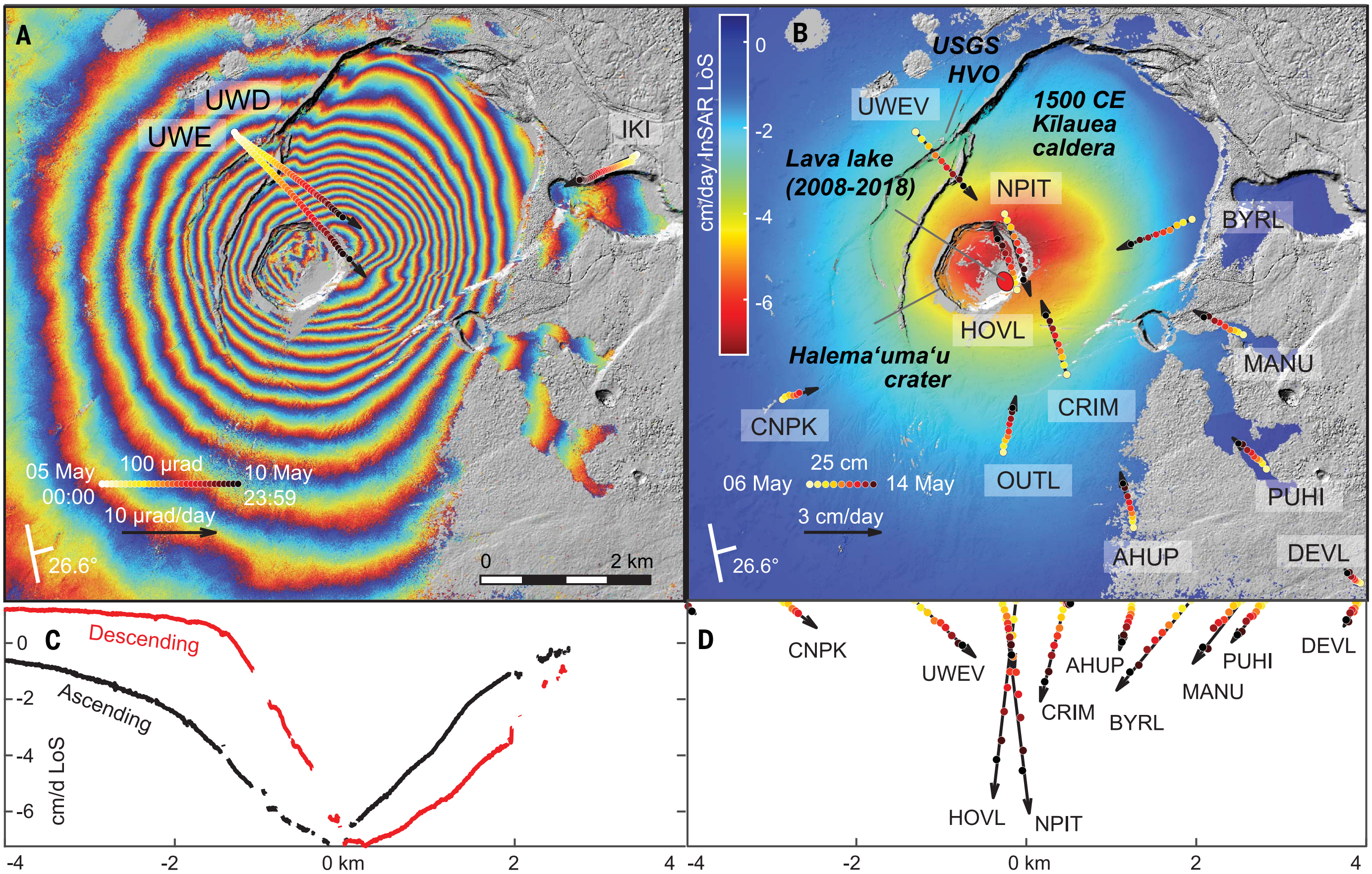


Fig. 2. Spatial pattern of subsidence at Kīlauea's summit in 2018. (**A**) Ground tilt overlaid on an ascending-mode COSMO-SkyMed interferogram spanning 6 to 10 May 2018 (table S1). Colored dots show observed tilt, and black arrows show best-fitting tilt velocities used for modeling. Each complete InSAR color fringe represents 1.55 cm of displacement in the look direction of the satellite (T symbol, 26.6° from vertical). Small-scale irregularities in the fringe pattern are evident in the caldera. Background shaded digital elevation model (DEM) shows Kīlauea's summit in 2009, similar to its appearance in April 2018. (**B**) Observed GPS displacements (colored dots) and best-fitting velocities (black arrows) overlaid on the unwrapped interferogram from (A). An active lava lake was nested within Halema'uma'u crater, itself nested in the larger 1500 CE Kīlauea caldera. LoS, line of sight. (**C**) West-east profiles of LoS COMSO-SkyMed InSAR velocities approximately through the center of Halema'uma'u crater. Profiles differ because of different look angles. (**D**) View of GPS data in (B), looking north.

modulus and magma density) to vary to account for their uncertainties. From the MCMC results and additional independent information we computed other parameters of interest, such as the rate of magma outflow from the reservoir. Parameter estimates take the form of probability density functions (PDFs), which account for uncertainties in data and prior information. We found that model output is consistent with the withdrawal rate of Kīlauea's lava lake and the first-order temporal and spatial pattern of ground deformation preceding caldera collapse (Fig. 6). We discuss our modeling results and implications throughout the following sections.

Location and geometry of subcaldera magma storage

Magma reservoir depth, volume, and geometry play a direct role in the onset, style, and duration of caldera collapse (*15*, *18*–*21*), but magma storage beneath most calderas is poorly understood and subject to controversy (*22*, *23*). Investigations at volcanoes that have hosted historic caldera-forming eruptions suggest that storage zones may be complex and occur over a range of depths (*3*, *24*–*28*). We found that geodetic data preceding caldera collapse at Kīlauea in 2018 are consistent with evacuation of magma from a storage reservoir centered at ~2 km depth just east of Halema'uma'u crater (Fig. 5 and table S2). The estimated magma reservoir is somewhat vertically elongated, as required to explain the observed ratio of vertical to horizontal displacements. The reservoir's depth implies an initial (pre-eruptive) magma pressure of ~45 MPa on the basis of the magmastatic lava lake relationship together with prior constraint on magma density (*8*). To the extent that magma density and lithostatic density were similar, the open lava lake vent precludes large magmatic overpressures before the onset of the eruption (*8*).

In the past two millennia, two long-lived, deep calderas have existed at the summit of Kīlauea: one from ~200 BCE to ~1000 CE, and the modern caldera, which formed in ~1500 CE and began refilling in ~1800 CE (*29*). Magma storage beneath Kīlauea's 1500 CE caldera was inferred in the first written records of the volcano nearly two centuries ago (*30*) and explains subsidence associated with rift zone intrusions and eruptions. At least two persistent magma reservoirs—the Halema'uma'u reservoir just east of Halem'uma'u crater and another at greater depth beneath the south part of the 1500 CE caldera—have been hypothesized on the basis of geodetic and other observations (*6*, *31*–*38*). Several transient storage zones may also have existed (*36*), and VLP seismic energy frequently emitted from a source ~1 km beneath the northeast rim of Halema'uma'u (*39*) has been interpreted as the intersection of north- and east-trending dikes (*11*, *40*). The geometries and relationships between these various magma storage regions have been difficult to interpret, and in some cases appear to change over time.

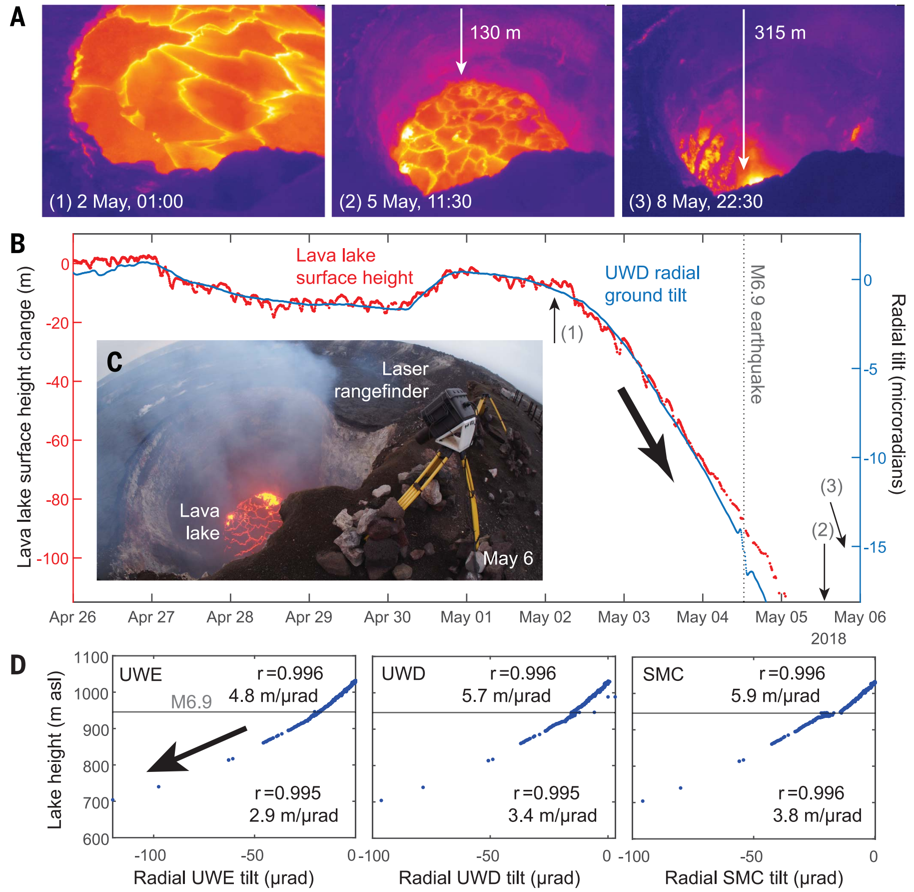


Fig. 3. Withdrawal of Kīlauea's lava lake in early May. (**A**) Thermal images of the lava lake surface taken from the south rim of Halemaʻumaʻu crater while the lake was draining. (**B**) Time series of change in lava lake surface height relative to 26 April, and radially outward low-pass–filtered ground tilt at UWD. Time series after 5 May are shown in Fig. 4. Numbers correspond to acquisition times of images in (A). (**C**) Photograph showing the lava lake on 6 May and the laser rangefinder used to measure its surface height. (**D**) Relationship between lava lake surface height and radially outward tilt (with M_w 6.9 earthquake offset approximately removed). At all stations, the ratio decreased by ~40% around the time of the M_w 6.9 earthquake, denoted by the horizontal gray line. Correlation coefficients are denoted by r.

PHOTO: U.S. GEOLOGICAL SURVEY

The reservoir location and geometry we estimate here lead us to conclude that magma withdrawal from the Halemaʻumaʻu reservoir was responsible for observed ground subsidence in 2018.

Misfits between model predictions and geodetic data provide additional insight into magma storage (Fig. 6). Our model closely fits lava lake withdrawal rate data but cannot account for small-scale features observed in the InSAR data (fig. S7), nor can it explain the very-high-quality GPS data to within formal uncertainties. Material heterogeneity such as preexisting faults and altered rocks, localized shallow magma storage, or irregularities in the top of the reservoir itself may be responsible for these features [we scale data uncertainties to account for these limitations (*8*)]. The model also inadequately accounts for subsidence observed south of the caldera. This likely reflects the early stages of magma drainage from Kīlauea's deeper and more enigmatic south caldera reservoir. Ground deformation believed to be due to magma evacuation from this reservoir increased in cumulative magnitude and spatial extent through June and July and continued after the cessation of the eruption (presumably as magma drained to refill the ERZ). However, most of the deformation during our modeled time period can be attributed to the Halemaʻumaʻu reservoir (predicted deformation from the model reduces variance in modeled InSAR scenes by 93 to 96%).

Volume of magma storage

The volume of magma stored beneath a volcano exerts a primary control on nearly all aspects of volcanic activity, including limiting the size of an eruption and any possible caldera collapse. Yet, magma storage volumes are very poorly known at almost all of Earth's volcanoes. Intensive study at Kīlauea over previous years has yielded estimates for the Halemaʻumaʻu reservoir varying over two orders of magnitude [from 0.2 to >20 km^3 (*6*, *34*, *41*–*44*)].

In general, geodetic data can be used to resolve the quantity $V\dot{p}/\mu$ for a magma reservoir, where V is reservoir volume, $\dot{p}$ is pressure change rate, and μ is host rock shear modulus, but not these terms independently (*45*). Our parameter estimation resolved V by using constraints on $\dot{p}$ from the lava lake data (below) and on μ from previous studies (*6*, *41*). Because $\dot{p}$ is much more tightly constrained than μ, we were able to resolve the ratio $V/\mu \cong 1.3 \pm 0.15$ m^3/Pa (*8*) (fig. S16). This implies that reservoir volume should be of the same order as the rigidity of the host rock. The combination of spatially dense geodetic data with the finite-source model used in our study provided additional constraint on reservoir volume (*45*), and the maximum size of the reservoir was geometrically limited by its depth and shape (both resolved geodetically).

We found that 2.5 to 7.2 km^3 of magma (at 68% credible bounds) was stored beneath the summit of the volcano in the Halemaʻumaʻu reservoir at the beginning of May 2018. The upper bound should be considered only approximate; volumes of 10 km^3 or even larger cannot strictly be ruled out by the data, particularly if we relax a priori limits on the presence of magma storage at very shallow depths (<750 m) (*8*). On the other hand, volumes of <1 km^3 are improbable, because smaller reservoirs cannot explain the high rate of observed ground deformation without requiring an unreasonably weak host rock (pressure change rate is tightly constrained by the lava lake data). Precollapse storage volumes for other basaltic calderas are not well known, but our calculated volume is far smaller than that of reservoirs inferred to have supplied large silicic caldera-forming eruptions.

Rate of magma depressurization and drainage

Reservoir pressure change rate $\dot{p}$ is constrained in our parameter estimation by the observed rate of lava lake withdrawal, the prior distribution on lava lake density, and the magmastatic assumption (*8*). Thus, $\dot{p}$ is insensitive to geodetic data and modeling. We estimated that pressure in the reservoir decreased at 1.25 ± 0.09 MPa/day (Fig. 5B) after the M_w 6.9 earthquake. At this rate, pressure at the reservoir's centroid would have decreased to atmospheric (an impossibility) by early June. Continuation of the eruption at a high rate for 3 months therefore required an increase of reservoir pressure through collapse of the overlying rock. This mechanism is also consistent with surges in effusion rate after collapses later in the eruption (*16*).

The volumetric rate of contraction $\dot{V}$ of the magma reservoir and the volumetric rate q

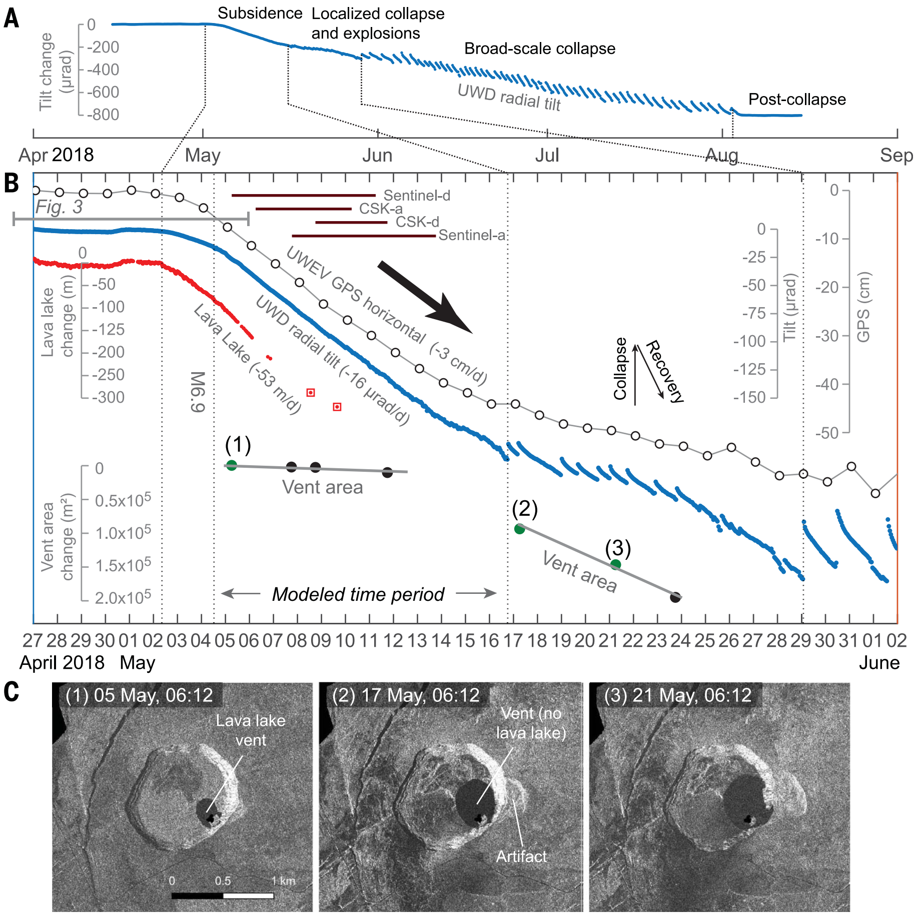


Fig. 4. Temporal evolution of summit deflation. (**A**) Radial ground tilt at UWD over the full eruption. Positive tilt is consistent with reservoir inflation (pressurization) and negative tilt with deflation. Collapses appear as small sawteeth from 16 to 26 May (nearly invisible at this scale) and as much larger sawteeth during broad-scale collapse (29 May and after). Time series were corrected for certain tectonic offsets. (**B**) GPS, tilt, lava lake surface height, and vent area time series indicating summit deflation from late April to early June 2018. Stations UWD (tilt) and UWEV (GPS) are approximately colocated (see Fig. 2 for station locations). Lava lake points with boxes were derived from structure-from-motion photogrammetry and are more uncertain. Vent area was inferred from satellite radar (ascending mode in green and descending mode in black) amplitude images as shown in (C); numbers on the time series correspond to these images. Time spans of modeled InSAR data are shown as horizontal bars and denoted with "-a" for ascending mode and "-d" for descending mode. The gray horizontal bar indicates the time span shown in Fig. 3. CSK, COSMO-SkyMed. (**C**) CSK radar amplitude images showing enlargement of the summit vent. Brighter pixels indicate higher radar reflectivity, so the vent appears black.

at which magma exited the reservoir are important to the timing of caldera collapse and the dynamics of summit draining and its relation with processes in the ERZ (*19*, *21*). We computed $\dot{V} = -1.3 \times 10^6 \pm 0.1\ \mathrm{m^3/day}$ ($\sim 15\ \mathrm{m^3/s}$) using estimated model parameters together with a numerical model for the elastic compressibility of the magma reservoir (*8*). This estimate is tightly constrained by the geodetic data. Combined with our posterior distribution for $\dot{p}$, we found that each pascal of pressure reduction in the reservoir reduced its volume by ~1 m^3 (dV/dp = 1.0 ± 0.1 m^3/Pa). Because of the rigidity of the host rock, the reservoir itself was contracting at only ~0.03% per day while its internal centroid pressure was decreasing at ~3% per day. At shallower depths in the reservoir, the relative pressure change rate would have been even greater.

Because magma is compressible, the rate at which the reservoir contracted was likely not equal to the rate of magma withdrawal. Using our distribution for $\dot{V}$ and independent constraint on compressibility (*8*), we estimated a net magma outflow rate q from the Halemaʻumaʻu reservoir of 2.3 million to 5.4 million m^3/day (27 to 62 m^3/s) at 68% credible bounds. This rate exceeds the average supply to Kīlauea from the mantle by an order of magnitude (*37*, *46*, *47*) and thus should approximate the total rate of flow to the ERZ from the contracting reservoir. Adding another ~5 to 10 m^3/s from the draining lava lake and its feeder conduit (*8*) yields a combined outflow rate of ~35 to 70 m^3/s from the lava lake and Halemaʻumaʻu reservoir. This is much higher than the time-averaged eruption rate from 3 to 18 May (7 m^3/s) (*48*), indicating that summit magma was entering the rift without erupting in order to feed deflation of the middle ERZ and growth of the LERZ intrusion. By June, after the onset of collapse events, LERZ eruption rates had increased by at least an order of magnitude (*7*), and the time-averaged rate of caldera collapse was ~two to five times larger than our estimated magma outflow rate. These observations strongly suggest a large increase in magma withdrawal rate from the summit in association with caldera collapse.

Reservoir failure thresholds

Placing bounds on the thresholds at which magma reservoirs begin to fail is important for determining the collapse hazard of an ongoing eruption (*49*), interpreting the geological record, and understanding the mechanical processes that lead to caldera collapse. Reservoir failure is triggered by stresses imparted to the host rock by changes in internal pressure. Kīlauea's lava lake provided a window into changing magma system pressure but disappeared from view ~1 week before the first collapse event. However, by assuming that pressure continued to decrease at rate $\dot{p}$ between the end of the modeled time period (14 May) and the first collapse (16 May), as suggested to first order by geodetic data, we estimated a pressure change at failure Δp_f = −17.2 ± 1.1 MPa (*8*).

We also used tilt data as a direct empirical proxy for pressure change, using the scaling relationship established while the lava lake was active (at UWD, 0.078 ± 0.006 MPa per microradian of radial tilt). This approach does not rely on any model except for the magmastatic relationship used to establish the scaling ratio, nor does it require an assumption of constant rates, but it can be affected by ground deformation caused by processes other than reservoir pressure change. We used this approach to estimate pressure changes after 16 May under the assumption that ground tilt during collapse events was caused entirely by changes in reservoir pressure [this likely overestimates pressure changes somewhat owing to faulting processes (*50*)]. With this approach, we obtained pressure changes of ~17.8 and ~25.0 MPa immediately before the first collapse event on 16 May (similar to the model-based results) and the first broad-scale collapse on 29 May, respectively (Fig. 7 and fig. S10) (*8*). These estimates

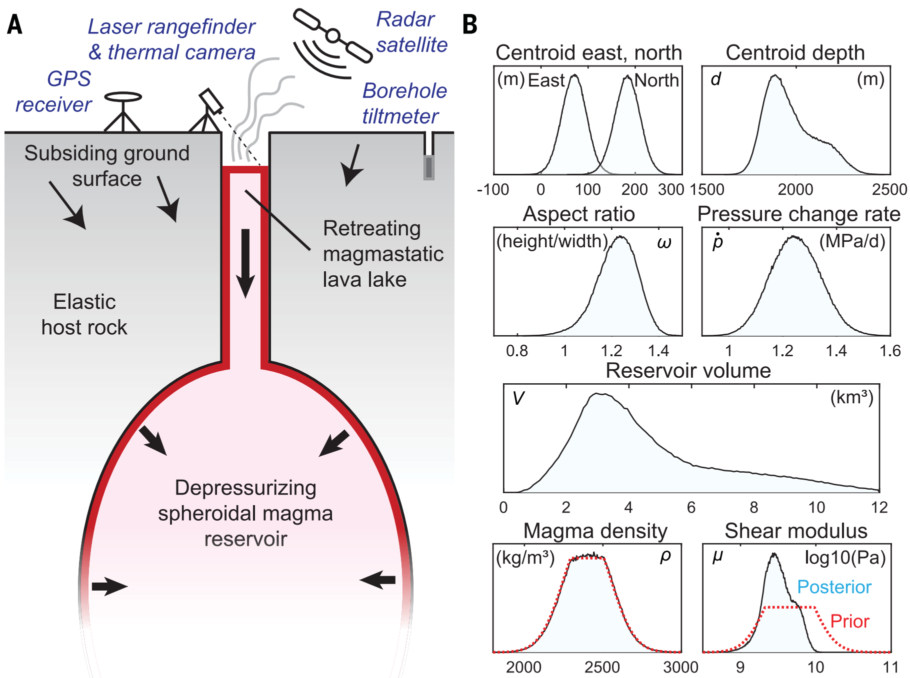


Fig. 5. Model geometry and estimated parameters. (**A**) Conceptual model geometry including instruments that recorded observations used in this study. The reservoir centroid is shown for simplicity directly beneath the lava lake, but this is not required in our model. (**B**) Marginal posterior PDFs of primary estimated model parameters (8), excluding "nuisance" parameters associated with data uncertainties (fig. S17). East and north positions are relative to 19.4073°N, 155.2784°W (the east rim of precollapse Halemaʻumaʻu crater), and depth is approximately relative to the volcano's summit.

imply a relative pressure reduction exceeding 30% at the reservoir's centroid by 16 May. They can also be related to shear stresses in the host rock, although the conditions required to trigger failure are complex and poorly understood. Using simple geometrical arguments, we computed the shear stress that the deflating reservoir imparted to an overlying cylindrical ring fault and estimated a stress change of between ~8 and 13 MPa (8, 18).

Although it is pressure changes that trigger collapse, due to the lack of observations at natural systems failure criteria are more typically formulated in terms of volume changes. Reservoir volume change may be tracked nearly in real-time using geodetic data, and erupted volume may be tracked directly or with geophysical observations. We defined critical fractions $V_{crit} = -\Delta V_f/V$ and $f_{crit} = -\Delta q_f/V$ (19), where ΔV_f and Δq_f are the reservoir volume change and total magma extraction volume at the time of failure, respectively. To estimate ΔV_f, we scaled the model-based estimate of Δp_f at the first collapse by the ratio dV/dp obtained from the Bayesian estimation results. Because $dV/dp \approx 1$, the magnitudes of pressure and volume changes were comparable. Scaling by reservoir volume yielded V_{crit} = 0.27 to 0.66%, and further scaling by system compressibility yielded f_{crit} = 0.68 to 2.2%, both at 68% credible bounds (table S2). At 95% confidence, we concluded that <3.5% of magma was evacuated before the onset of collapse at Kīlauea.

Geometry of the roof block

The aspect ratio of the roof block above a magma reservoir (Fig. 8, C and D) influences not only the timing of collapse onset but also its subsequent structural development and style (20, 51, 52). In general, low-aspect roof blocks [$R_a < 1$, where R_a is the thickness T of the crust above the magma reservoir divided by the reservoir diameter D (52)] tend to favor a central coherent collapse "piston" bounded by reverse faults, whereas high-aspect ($R_a > 1$) blocks favor incoherent subsidence through migration of fractures upward from the reservoir. However, observational constraints on R_a from real-world caldera collapses are limited, owing to poor knowledge of the geometry of subcaldera magma reservoirs. Caldera diameter must generally be used as a proxy for reservoir diameter and roof thickness inferred roughly from geological or geophysical data (18, 19, 53, 54).

The set of finite-source geodetic models derived from our MCMC analysis allowed us to estimate R_a. Taking T to be the distance between the surface and the top depth of each magma reservoir in the posterior probability distribution, we found that the roof block at Kīlauea was thin and wide, with $R_a \approx$ 0.4 (Fig. 8). R_a would be smaller if we were to relax our minimum reservoir top depth (8) but would be larger if we measured height from a point other than its very top. Small reservoirs from our probability distribution yield aspect ratios closer to 1, but in general $R_a > 1$ appears unlikely.

Reservoir evacuation and the end of the eruption

It is often assumed that caldera-forming eruptions are terminated by the near-complete evacuation of their source reservoirs (3, 49, 54, 55), as suggested by some models (56) and perhaps indicated by long repose periods after some collapses (55). This hypothesis has implications for hazards during ongoing eruptions. It also allows for interpreting data from past events because it implies that erupted volume is approximately equal to reservoir volume. Although there is evidence that this assumption may not be valid (20, 56), it has been difficult to evaluate because of limited knowledge of subcaldera magma reservoir volumes.

Taking the total 2018 summit collapse volume (7) as a proxy for the total volume change of the shallow reservoir during the eruption, we used our posterior PDF for reservoir volume to estimate that only 11 to 33% of Kīlauea's shallow magma reservoir was evacuated by the end of the eruption. The probability of complete drainage is very small; we estimated <5% probability that even half of the reservoir was drained (Fig. 8). This inference is consistent with the relative constancy of collapse-related geophysical signals from June to August (7), which might have changed in character if the reservoir had neared complete evacuation, and also with the post-eruptive return of episodic days-long ground deformation cycles at the summit, which are believed to be caused by pressure perturbations in the shallow magma reservoir (6). Our results suggest caution in assuming that magma reservoirs (at least basaltic ones) fully evacuate during caldera-forming eruptions.

Discussion

Caldera collapse at Kīlauea in 2018 was caused by high-rate magma evacuation from a roughly equant storage zone of several cubic kilometers at shallow depth (~2 km), centered just east of the former Halemaʻumaʻu crater. Many previous studies have inferred magma storage in this area, but 2018 data provide new insights. Our simple geodetic model cannot account for magma withdrawal from other reservoirs or the fine-scale topology of magma storage [for instance, we likely cannot rule out magma stored in a broad plexus of interconnected magma-filled cracks (57) with similar magma volume], but it well explains the observed overall spatial pattern of ground deformation. Likewise, the rate of magma system depressurization estimated by our model can

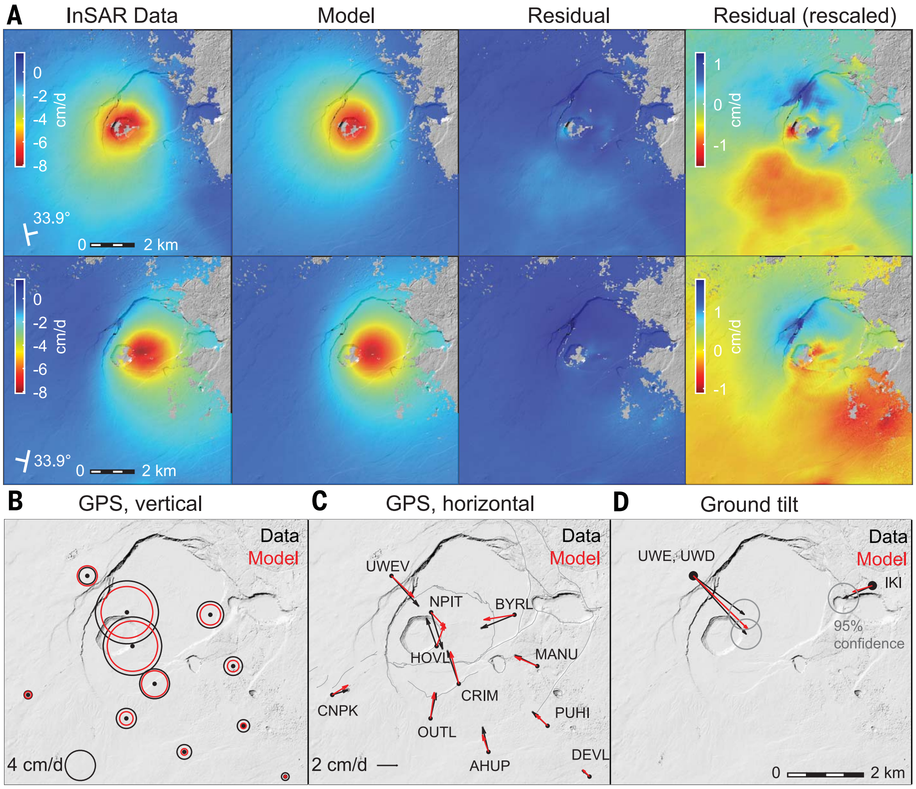


Fig. 6. Fit of model to observations. Shown are predictions from the mean of the posterior distribution. We do not show lava lake data, which the model is able to fit "exactly" (to within an arbitrary precision). (**A**) Sentinel-1 ascending- and descending-mode interferograms (see fig. S15 for COSMO-SkyMed). The variance of the InSAR data is reduced by more than 95% after subtracting model predictions. Residuals in and south of the caldera do remain (the images in the rightmost column have a different color scale to highlight these effects). (**B**) Vertical GPS velocities. (**C**) Horizontal GPS velocities. Formal 95% data uncertainty ellipses are shown but are too small to be easily visible; in the estimation, these uncertainties are scaled using data-weighting hyperparameters (*8*). (**D**) Ground tilt rates.

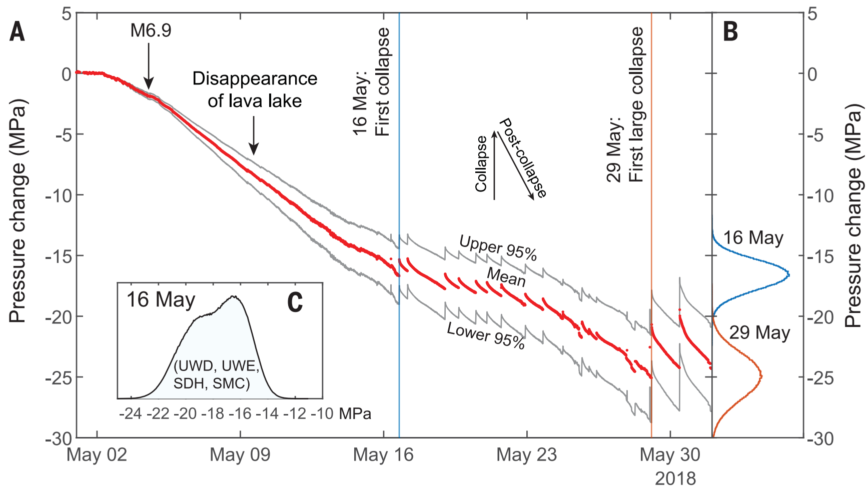


Fig. 7. Pressure change in the magma reservoir. (**A**) Time series of reservoir pressure change derived from scaled tilt at UWD. The time span is similar to that in Fig. 4. Uncertainties are due to lava lake density and the lake-tilt ratio (Fig. 3). Certain offsets not apparently related to magmatic processes were removed from UWD tilt data. (**B**) Marginal distributions for pressure change immediately preceding the first collapse (16 May) and the first large collapse (29 May). We combined marginal distributions for tiltmeters UWD, UWE, SDH, and SMC to produce the distribution in (**C**).

explain the observed rate of lava lake withdrawal and ground deformation.

When did caldera collapse begin? Seismicity after the M_w 6.9 earthquake might have indicated the early stages of caldera-fault propagation at depth (*58*), but there appeared to be relatively little effect on surface deformation during the first half of May (*8*), and there was no geophysical evidence for collapse of rock into the deeper magmatic system. Quasi-periodic VLP seismic and geodetic signals recorded from 16 to 26 May were associated with vent widening, volume loss, and ejection of ash, but not surface faulting over a broad area. Yet, InSAR data from this time showed a more complex deformation pattern in the caldera than that present earlier in the month, suggestive of the early-stage surface expression of slip on buried caldera faults. Furthermore, geophysical signals were similar to those recorded during caldera collapses at other volcanoes and at Kīlauea after 29 May, when broadscale collapse was visually observed. Thus, the events of 16 to 26 May were evidently related to collapse of rock into the magmatic system, although the extent to which these collapses occurred into the lava lake feeder conduit and/or shallow dike-like storage bodies, as opposed to the Halemaʻumaʻu reservoir, remains an open question. Also unclear is the extent to which any propagation of buried caldera faults during this time related to geophysical observations. Nonetheless, we conclude that caldera collapse effectively began on 16 May, accelerated and enlarged on 29 May (when we were able to closely tie visual observations of broader-scale collapse to geophysical signals), and did not reach its full surface expression until late June.

The critical thresholds required for caldera collapse are thought to be controlled by many factors, including the shape (aspect ratio) of the roof rock above the reservoir (*18*, *19*); exsolved magmatic volatiles, which buffer pressure drop due to magma extraction (*56*, *59*, *60*); and preexisting faults and weaknesses (*49*). At Kīlauea, the 2018 collapse occurred within an older, larger caldera and, in some areas, appeared to proceed along preexisting faults. We speculate that both the empty lava lake vent and the relatively thin and wide roof block might have promoted failure (*18*, *19*). It is also possible that, at shallow depths, the retreating magma surface could have encountered a flared conduit geometry, leading to instability. An open question is how critical failure thresholds might differ between small nested-caldera basaltic systems, such as Kīlauea, and large silicic systems.

Caldera collapse began at Kīlauea after the elastic reservoir had contracted only very slightly (V_{crit} < 1.1%), caused by withdrawal of only a very small fraction of its stored magma

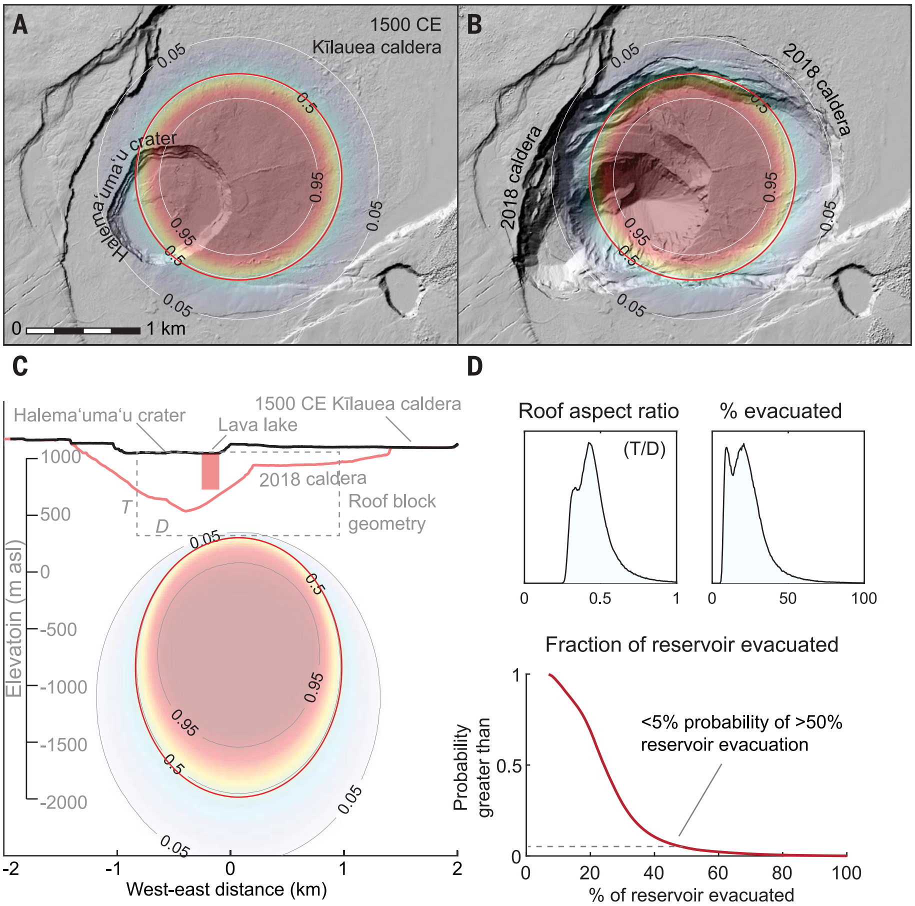


Fig. 8. Probabilistic magma storage in the Halema'uma'u reservoir beneath Kīlauea's summit. Contours and shading indicate estimated probability of magma storage based on the range of model geometries inferred in the parameter estimation (8). (**A** and **B**) Results for a horizontal slice near the reservoir centroid at 2 km depth. (**C**) Probability along an east-west slice at the reservoir centroid. Model depths are converted to vertical elevations using the approximate mean geodetic observation elevation [1100 m above sea level (asl)]. Colors indicate relative probability (red, more likely; blue and white, less likely). Red circles show geometry predicted by the median of the posterior distribution. Shaded DEMs in (A) and (B) show the summit as it appeared before and after the 2018 caldera collapse, respectively. The dashed rectangle above the storage zone in (C) shows the rough geometry of the roof block. The bulk of magma was stored below sea level and the subaerial ERZ vents (Fig. 1). (**D**) Posterior PDFs of roof aspect ratio and the probability of complete reservoir evacuation given the observed caldera collapse volume, along with complementary cumulative distribution.

(f_{crit} < 4%). Geological observations and models have suggested that f_{crit} may range from <10% to >90% (18, 19), but direct evidence has been lacking (note that many studies do not distinguish between V_{crit} and f_{crit}, which are equal only if magma is incompressible). Geophysical observations from basaltic collapses at Piton de la Fournaise, Fernandina (Galápagos), Miyakejima, and Bárðarbunga volcanoes yielded f_{crit} of ~8 to 20%, in some cases much lower than values suggested by analog models (3, 49) but still much higher than we found for Kīlauea. Although it is possible that collapse began unusually quickly at Kīlauea, these previous estimates had to rely on assumptions that the volumes of initial collapse events were comparable to precollapse magma withdrawal volumes and that eruptions completely drained their magma reservoirs (3, 49, 54). As we have shown here, these assumptions are not always valid and could lead to a substantial overestimation of f_{crit}. These discrepancies indicate that calderas may fail more quickly than previously understood.

Although it is changes in magma pressure that drive host rock failure and caldera collapse, robust estimates of precollapse pressure changes have previously been unavailable. Magma extraction volumes are far more readily measured in nature but are only relevant to collapse to the extent that they influence reservoir pressure (an effect modulated by the compressibility of magma in the reservoir). Data from Kīlauea allowed us to move beyond reliance on f_{crit} and directly estimate precollapse pressure change. Knowledge of the pressure change makes it possible to compute stress changes on the roof block and thus tie the observations to the failure process.

Once failure began, episodic roof block collapse transferred the load of the overlying rock to the magma, increasing its pressure. This process may explain similar episodic geophysical observations at other basaltic caldera collapses (14, 15, 61). By using ground tilt as a proxy for reservoir pressure change, we estimated that inflationary deformation during the first collapse event on 16 May was caused by a pressure increase of ~1.3 MPa in the reservoir, only a fraction of the preceding deflation. Because reservoir pressure was likely near lithostatic at the onset of the eruption, this result indicates incomplete repressurization of the reservoir after the onset of collapse and implies residual frictional strength on the walls of the collapsing block(s) such that the weight of the roof was not entirely supported by the magma. This finding stands in contrast to assumptions that roof collapses reestablish lithostatic pressure in the reservoir (56, 59) but supports the results of some numerical models (62).

The surface expression of caldera collapse was complex, asymmetric, and evolving, consisting of funnel-like gravitational failure into the evacuated lava lake vent and piston-like slumping of coherent blocks as large as ~150 ha, in some cases clearly bounded by preexisting faults. Taken as a whole, these events were consistent with collapse of roof rock into a shallow reservoir, governed not only by the aspect ratio of the roof but also by preexisting caldera faults and structural weaknesses, and possibly shallow unmodeled magma storage [e.g., (11, 63)]. These observations are consistent with geological investigations and numerical experiments that demonstrate the complex diversity of collapse styles that can occur during caldera formation (51, 64).

The location and lateral extent of magma storage inferred from our model are similar to the final geometry of the 2018 caldera collapse (Fig. 8). To first order, the relationship between the range of plausible reservoir geometries and observed caldera dimensions favors primary collapse faults ranging from near-vertical to inward dipping. Results indicate that the shallow subcaldera magma storage system spanned only a portion of the caldera in existence from 1500 CE to the present. The larger magma storage body required to explain the 1500 CE collapse may have been partially destroyed then or in a subsequent event (such as a large collapse that occurred at the volcano in 1868) or may

have involved failure of deeper parts of the summit magma system.

Globally, lava lakes are rare. Where they do exist, close observation during magma draining events may bear rich dividends, particularly if relayed in real time. Some of the data used in this study were evaluated in rapid-response mode internally by the USGS during the eruption with a preliminary form of our model. Resulting parameter estimates were used to better understand the possible course of the eruption and guided our thinking about hazards as the eruption progressed, highlighting the importance of near–real-time data and modeling capabilities at the world's volcano observatories.

Outlook

Despite insights into volcanic calderas afforded over the past two decades by well-documented collapses at Miyakejima, Piton de la Fournaise, and Bárðarbunga volcanoes, the conditions that trigger the onset of collapse remain only poorly understood. Draining of Kīlauea's summit lava lake in 2018 yielded a window into changing pressure in the volcano's shallow magma reservoir. We tracked the evolution of the magmatic system as it underwent steady high-rate elastic decompression due to magma withdrawal, followed by episodic fault-bounded caldera collapse. We were able to quantify the changing pressure in the reservoir, which, together with geodetic data, made it possible to estimate the volume of magma storage and the critical thresholds that preceded the onset of collapse. Caldera collapse began due to a relatively large decrease in the magma reservoir's internal pressure caused by withdrawal of only a small fraction of stored magma. Episodic fault-bounded subsidence of the roof block above the reservoir increased magma pressure, sustaining the flow of magma and thus representing a critical turning point in the evolution of the eruption.

REFERENCES AND NOTES

1. H. Williams, in *Bulletin of the Department of Geological Sciences*, G. D. Louderback, C. A. Anderson, C. L. Camp, R. W. Chaney, H. Williams, Eds. (Univ. of California Publications, ed. 25, 1941), pp. 239–346.
2. A. Geyer, J. Martí, The new worldwide collapse caldera database (CCDB): A tool for studying and understanding caldera processes. *J. Volcanol. Geotherm. Res.* **175**, 334–354 (2008). doi: 10.1016/j.jvolgeores.2008.03.017
3. M. T. Gudmundsson *et al*., Gradual caldera collapse at Bárdarbunga volcano, Iceland, regulated by lateral magma outflow. *Science* **353**, aaf8988 (2016). doi: 10.1126/science.aaf8988; pmid: 27418515
4. M. R. Patrick, K. R. Anderson, M. P. Poland, T. R. Orr, D. A. Swanson, Lava lake level as a gauge of magma reservoir pressure and eruptive hazard. *Geology* **43**, 831–834 (2015). doi: 10.1130/G36896.1
5. M. Patrick, D. Swanson, T. Orr, A review of controls on lava lake level: Insights from Halema'uma'u Crater, Kīlauea Volcano. *Bull. Volcanol.* **81**, 13 (2019). doi: 10.1007/s00445-019-1268-y
6. K. R. Anderson, M. P. Poland, J. H. Johnson, A. Miklius, "Episodic deflation-inflation events at Kīlauea Volcano and implications for the shallow magma system," chapter 11 in *Hawaiian Volcanism: From Source to Surface*, R. J. Carey, M. P. Poland, V. Cayol, D. Weis, Eds. (AGU Geophysical Monograph Series vol. 208, AGU, 2015), pp. 229–250; doi: 10.1002/9781118872079.ch11.
7. C. A. Neal *et al*., The 2018 rift eruption and summit collapse of Kīlauea Volcano. *Science* **363**, 367–374 (2019). doi: 10.1126/science.aav7046; pmid: 30538164
8. Materials and methods are available as supplementary materials.
9. M. R. Patrick, E. F. Younger, W. Tollett, Lava level and crater geometry data during the 2018 lava lake draining at Kilauea Volcano, Hawaii, ScienceBase (2019). doi: 10.5066/P9MJY24N
10. H. L. Kehoe, E. D. Kiser, P. G. Okubo, The rupture process of the 2018 Mw 6.9 Hawai'i earthquake as imaged by a genetic algorithm-based back-projection technique. *Geophys. Res. Lett.* **46**, 2467–2474 (2019). doi: 10.1029/2018GL080397
11. P. Dawson, B. Chouet, Characterization of very-long-period seismicity accompanying summit activity at Kīlauea Volcano, Hawai'i: 2007–2013. *J. Volcanol. Geotherm. Res.* **278–279**, 59–85 (2014). doi: 10.1016/j.jvolgeores.2014.04.010
12. H. Munekane, J. Oikawa, T. Kobayashi, Mechanisms of step-like tilt changes and very long period seismic signals during the 2000 Miyakejima eruption: Insights from kinematic GPS. *J. Geophys. Res.* **121**, 2932–2946 (2016). doi: 10.1002/2016JB012795
13. T. Staudacher *et al*., The April 2007 eruption and the Dolomieu crater collapse, two major events at Piton de la Fournaise (La Réunion Island, Indian Ocean). *J. Volcanol. Geotherm. Res.* **184**, 126–137 (2009). doi: 10.1016/j.jvolgeores.2008.11.005
14. L. Michon, N. Villeneuve, T. Catry, O. Merle, How summit calderas collapse on basaltic volcanoes: New insights from the April 2007 caldera collapse of Piton de la Fournaise volcano. *J. Volcanol. Geotherm. Res.* **184**, 138–151 (2009). doi: 10.1016/j.jvolgeores.2008.11.003
15. H. Kumagai *et al*., Very-long-period seismic signals and caldera formation at Miyake Island, Japan. *Science* **293**, 687–690 (2001). doi: 10.1126/science.1062136; pmid: 11474109
16. M. R. Patrick *et al*., Cyclic lava effusion during the 2018 eruption of Kīlauea Volcano. *Science* **366**, eaay9070 (2019). doi: 10.1126/science.aay9070
17. M. Gu, J. O. Berger, Parallel partial Gaussian process emulation for computer models with massive output. *Ann. Appl. Stat.* **10**, 1317–1347 (2016). doi: 10.1214/16-AOAS934
18. O. Roche, T. H. Druitt, Onset of caldera collapse during ignimbrite eruptions. *Earth Planet. Sci. Lett.* **191**, 191–202 (2001). doi: 10.1016/S0012-821X(01)00428-9
19. A. Geyer, A. Folch, J. Martí, Relationship between caldera collapse and magma chamber withdrawal: An experimental approach. *J. Volcanol. Geotherm. Res.* **157**, 375–386 (2006). doi: 10.1016/j.jvolgeores.2006.05.001
20. N. Geshi, J. Ruch, V. Acocella, Evaluating volumes for magma chambers and magma withdrawn for caldera collapse. *Earth Planet. Sci. Lett.* **396**, 107–115 (2014). doi: 10.1016/j.epsl.2014.03.059
21. S. Kusumoto, A. Gudmundsson, Magma-chamber volume changes associated with ring-fault initiation using a finite-sphere model: Application to the Aira caldera, Japan. *Tectonophysics* **471**, 58–66 (2009). doi: 10.1016/j.tecto.2008.09.001
22. B. M. Kennedy *et al*., Magma plumbing beneath collapse caldera volcanic systems. *Earth Sci. Rev.* **177**, 404–424 (2018). doi: 10.1016/j.earscirev.2017.12.002
23. J. Martí, A. Geyer, A. Folch, J. Gottsmann, in *Caldera Volcanism: Analysis Modelling and Response* (Elsevier, 2008), vol. 10, pp. 233–283.
24. A. Di Muro *et al*., The shallow plumbing system of Piton de la Fournaise Volcano (La Réunion Island, Indian Ocean) revealed by the major 2007 caldera-forming eruption. *J. Petrol.* **55**, 1287–1315 (2014). doi: 10.1093/petrology/egu025
25. G. Saito, Y. Morishita, H. Shinohara, Magma plumbing system of the 2000 eruption of Miyakejima volcano, Japan, deduced from volatile and major component contents of olivine-hosted melt inclusions. *J. Geophys. Res.* **115**, B11202 (2010). doi: 10.1029/2010JB007433
26. M. Bagnardi, F. Amelung, Space-geodetic evidence for multiple magma reservoirs and subvolcanic lateral intrusions at Fernandina Volcano, Galápagos Islands. *J. Geophys. Res.* **117**, 1–19 (2012). doi: 10.1029/2012JB009465
27. A. Belousov, M. Belousova, B. Edwards, A. Volynets, D. Melnikov, Overview of the precursors and dynamics of the 2012–13 basaltic fissure eruption of Tolbachik Volcano, Kamchatka, Russia. *J. Volcanol. Geotherm. Res.* **307**, 22–37 (2015). doi: 10.1016/j.jvolgeores.2015.06.013
28. W. W. Chadwick *et al*., The May 2005 eruption of Fernandina volcano, Galápagos: The first circumferential dike intrusion observed by GPS and InSAR. *Bull. Volcanol.* **73**, 679–697 (2011). doi: 10.1007/s00445-010-0433-0
29. D. A. Swanson *et al*., Cycles of explosive and effusive eruptions at Kīlauea Volcano, Hawai'i. *Geology* **42**, 631–634 (2014). doi: 10.1130/G35701.1
30. W. Ellis, *Narrative of a Tour Through Hawaii, Or Owhyhee* (Fisher and Jackson, 1825).
31. S. Baker, F. Amelung, Top-down inflation and deflation at the summit of Kīlauea Volcano, Hawaii observed with InSAR. *J. Geophys. Res.* **117**, B12406 (2012). doi: 10.1029/2011JB009123
32. R. S. Fiske, W. T. Kinoshita, Inflation of Kilauea volcano prior to its 1967–1968 eruption. *Science* **165**, 341–349 (1969). doi: 10.1126/science.165.3891.341; pmid: 17809512
33. P. F. Cervelli, A. Miklius, in *The Pu'u 'O'o-Kupaianaha Eruption of Kilauea Volcano, Hawai'i: The First 20 Years*, C. Heliker, D. Swanson, J. T. Takahashi, Eds. (USGS Professional Paper 1676, 2003), pp. 149–164.
34. M. P. Poland, J. Sutton, T. M. Gerlach, Magma degassing triggered by static decompression at Kilauea Volcano, Hawai'i. *Geophys. Res. Lett.* **36**, L16306 (2009). doi: 10.1029/2009GL039214
35. P. Lundgren *et al*., Evolution of dike opening during the March 2011 Kamoamoa fissure eruption, Kīlauea Volcano, Hawai'i. *J. Geophys. Res.* **118**, 897–914 (2013). doi: 10.1002/jgrb.50108
36. M. P. Poland, A. Miklius, E. K. Montgomery-Brown, Magma supply, storage, and transport at shield-stage Hawaiian volcanoes. *U.S. Geol. Surv. Prof. Pap. 1801*. **2010**, 1–52 (2014).
37. K. R. Anderson, M. P. Poland, Bayesian estimation of magma supply, storage, and eruption rates using a multiphysical volcano model: Kīlauea Volcano, 2000–2012. *Earth Planet. Sci. Lett.* **447**, 161–171 (2016). doi: 10.1016/j.epsl.2016.04.029
38. J. Dvorak, A. Okamura, J. H. Dieterich, Analysis of surface deformation data, Kilauea Volcano, Hawaii: October 1966 to September 1970. *J. Geophys. Res.* **88**, 9295–9304 (1983). doi: 10.1029/JB088iB11p09295
39. T. Ohminato, B. A. Chouet, P. Dawson, S. Kedar, Waveform inversion of very long period impulsive signals associated with magmatic injection beneath Kilauea volcano, Hawaii. *J. Geophys. Res.* **103**, 23839–23862 (1998). doi: 10.1029/98JB01122
40. B. Chouet, P. Dawson, Seismic source dynamics of gas-piston activity at Kīlauea Volcano, Hawai'i. *J. Geophys. Res.* **120**, 2525–2560 (2015). doi: 10.1002/2014JB011789
41. P. Segall, P. Cervelli, S. Owen, M. Lisowski, A. Miklius, Constraints on dike propagation from continuous GPS measurements. *J. Geophys. Res.* **106**, 19301–19317 (2001). doi: 10.1029/2001JB000229
42. D. J. Johnson, Dynamics of magma storage in the summit reservoir of Kilauea Volcano, Hawaii. *J. Geophys. Res.* **97**, 1807–1820 (1992). doi: 10.1029/91JB02839
43. A. J. Pietruszka, D. E. Heaton, J. P. Marske, M. O. Garcia, Two magma bodies beneath the summit of Kīlauea Volcano unveiled by isotopically distinct melt deliveries from the mantle. *Earth Planet. Sci. Lett.* **413**, 90–100 (2015). doi: 10.1016/j.epsl.2014.12.040
44. A. J. Pietruszka, J. P. Marske, D. E. Heaton, M. O. Garcia, J. M. Rhodes, An isotopic perspective into the magmatic evolution and architecture of the rift zones of Kīlauea Volcano. *J. Petrol.* **59**, 2311–2352 (2018). doi: 10.1093/petrology/egy098
45. D. F. McTigue, Elastic stress and deformation near a finite spherical magma body: Resolution of the point source paradox. *J. Geophys. Res.* **92**, 12931 (1987). doi: 10.1029/JB092iB12p12931
46. T. L. Wright, F. W. Klein, in *Dynamics of Crustal Magma Transfer, Storage and Differentiation*, C. Annen, G. F. Zellmer, Eds. (Geological Society of London, 2008), vol. 304, pp. 83–116. doi: 10.1144/SP304.5
47. D. Dzurisin, M. P. Poland, in *Field Volcanology: A Tribute to the Distinguished Career of Don Swanson*, M. P. Poland, M. O. Garcia, V. E. Cam, A. Grunder, Eds. (Geological Society of America, 2018), vol. 2538, pp. 275–295. doi: 10.1130/2018.2538(12)
48. P. R. Lundgren, M. Bagnardi, H. Dietterich, Topographic changes during the 2018 Kīlauea eruption from single-pass airborne InSAR. *Geophys. Res. Lett.* **46**, 9554–9562 (2019). doi: 10.1029/2019GL083501
49. L. Michon, F. Massin, V. Famin, V. Ferrazzini, G. Roult, Basaltic calderas: Collapse dynamics, edifice deformation, and variations of magma withdrawal. *J. Geophys. Res.* **116**, B03209 (2011). doi: 10.1029/2010JB007636
50. P. Segall, K. R. Anderson, I. Johanson, A. Miklius, Mechanics of inflationary deformation during caldera collapse: Evidence from the 2018 Kīlauea eruption. *Geophys. Res. Lett.* 2019GL084689 (2019). doi: 10.1029/2019GL084689

51. E. P. Holohan, M. P. J. Schöpfer, J. J. Walsh, Mechanical and geometric controls on the structural evolution of pit crater and caldera subsidence. *J. Geophys. Res.* **116**, B07202 (2011). doi: 10.1029/2010JB008032
52. O. Roche, T. H. Druitt, O. Merle, Experimental study of caldera formation. *J. Geophys. Res. Solid Earth* **105**, 395–416 (2000). doi: 10.1029/1999JB900298
53. R. Scandone, V. Acocella, Control of the aspect ratio of the chamber roof on caldera formation during silicic eruptions. *Geophys. Res. Lett.* **34**, L22307 (2007). doi: 10.1029/2007GL032059
54. J. Stix, T. Kobayashi, Magma dynamics and collapse mechanisms during four historic caldera-forming events. *J. Geophys. Res.* **113**, B09205 (2008). doi: 10.1029/2007JB005073
55. A. Folch, J. Martí, Time-dependent chamber and vent conditions during explosive caldera-forming eruptions. *Earth Planet. Sci. Lett.* **280**, 246–253 (2009). doi: 10.1016/j.epsl.2009.01.035
56. J. Martí, A. Folch, A. Neri, M. Giovanni, Pressure evolution during explosive caldera-forming eruptions. *Earth Planet. Sci. Lett.* **175**, 275–287 (2000). doi: 10.1016/S0012-821X(99)00296-4
57. B. A. Chouet, P. B. Dawson, M. R. James, S. J. Lane, Seismic source mechanism of degassing bursts at Kilauea Volcano, Hawaii: Results from waveform inversion in the 10–50 s band. *J. Geophys. Res.* **115**, B09311 (2010). doi: 10.1029/2009JB006661
58. D. R. Shelly, W. Thelen, P. Okubo, "Anatomy of a caldera collapse: Kilauea 2018 summit seismicity sequence in high resolution," presented at the Seismological Society of America Annual Meeting, Seattle, WA, 23 to 26 April 2019.
59. T. H. Druitt, R. S. J. Sparks, On the formation of calderas during ignimbrite eruptions. *Nature* **310**, 679–681 (1984). doi: 10.1038/310679a0
60. S. M. Bower, A. W. Woods, Control of magma volatile content and chamber depth on the mass erupted during explosive volcanic eruptions. *J. Geophys. Res.* **102**, 10273–10290 (1997). doi: 10.1029/96JB03176
61. T. Simkin, K. A. Howard, Caldera collapse in the Galápagos Islands, 1968. *Science* **169**, 429–437 (1970). doi: 10.1126/science.169.3944.429; pmid: 17739001
62. E. P. Holohan, M. P. J. Schöpfer, J. J. Walsh, Stress evolution during caldera collapse. *Earth Planet. Sci. Lett.* **421**, 139–151 (2015). doi: 10.1016/j.epsl.2015.03.003
63. L. Gailler *et al*., 3D electrical conductivity imaging of Halema'uma'u lava lake (Kīlauea volcano). *J. Volcanol. Geotherm. Res.* **381**, 185–192 (2019). doi: 10.1016/j.jvolgeores.2019.06.001
64. P. W. Lipman, Subsidence of ash-flow calderas: Relation to caldera size and magma-chamber geometry. *Bull. Volcanol.* **59**, 198–218 (1997). doi: 10.1007/s004450050186
65. European Space Agency Sentinel Data Access; https://sentinel.esa.int/web/sentinel/sentinel-data-access.
66. Hawaiian Volcanoes Supersite; http://geo-gsnl.org/supersites/permanent-supersites/hawaiian-volcanoes-supersite/.
67. National Center for Airborne Laser Mapping, Hawaii Big Island Survey, OpenTopography (2012); doi: 10.5069/G9DZ067X
68. Kīlauea LiDAR Data (2018); https://kilauealidar.com.
69. I. A. Johanson, A. Miklius, Tiltmeter data from Kīlauea Volcano, Hawaii, spanning the 2018 eruption and earthquake sequence. U.S. Geological Survey data release, ScienceBase (2019). doi: 10.5066/P9310M9N
70. UNAVCO; https://unavco.org.

ACKNOWLEDGMENTS

E. Rumpf analyzed vent collapse from HVO webcam photos. P. Cervelli assisted with implementation of the analytical deformation model. M. McLay and Y. Zheng assisted with interferogram processing. This work benefited from numerous discussions with scientists at the Hawaiian Volcano Observatory and throughout the USGS. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. **Funding:** This work was funded by the USGS Volcano Hazards Program. **Author contributions:** K.R.A. conceptualized the project, analyzed data, developed the model, performed inversions, and coordinated manuscript writing. I.A.J. operated geodetic instruments, analyzed geodetic data, and contributed to modeling. M.R.P. installed and operated lava lake instrumentation and analyzed lava lake data. M.G. implemented the emulator, analyzed data uncertainties, and contributed to the Bayesian inversion. P.S. contributed to conceptualization, modeling, and validation of results. M.P.P. processed and analyzed InSAR data. E.K.M.-B. interpreted results and contributed to modeling. A.M. and all other authors contributed to data interpretation and manuscript production, and all USGS authors contributed to the eruption response and data collection. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sentinel SAR data are available from (*65*); COSMO-SkyMed SAR data from (*66*); DEM data from (*67*, *68*), tilt data from (*69*), GPS data from (*70*), and lava lake data from (*9*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaaz1822/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 and S2
References (*71*–*117*)
Movie S1

20 August 2019; accepted 13 November 2019
10.1126/science.aaz1822

RESEARCH ARTICLE SUMMARY

HIV VACCINES

Targeted selection of HIV-specific antibody mutations by engineering B cell maturation

Kevin O. Saunders*†, Kevin Wiehe*, Ming Tian*, Priyamvada Acharya*, Todd Bradley, S. Munir Alam, Eden P. Go, Richard Scearce, Laura Sutherland, Rory Henderson, Allen L. Hsu, Mario J. Borgnia, Haiyan Chen, Xiaozhi Lu, Nelson R. Wu, Brian Watts, Chuancang Jiang, David Easterhoff, Hwei-Ling Cheng, Kelly McGovern, Peyton Waddicor, Aimee Chapdelaine-Williams, Amanda Eaton, Jinsong Zhang, Wes Rountree, Laurent Verkoczy, Mark Tomai, Mark G. Lewis, Heather R. Desaire, Robert J. Edwards, Derek W. Cain, Mattia Bonsignori, David Montefiori, Frederick W. Alt†, Barton F. Haynes†

INTRODUCTION: A major goal of HIV-1 vaccine development is the design of immunogens that induce broadly neutralizing antibodies (bnAbs). However, vaccination of humans has not resulted in the induction of affinity-matured and potent HIV-1 bnAbs. To devise effective vaccine strategies, we previously determined the maturation pathway of select HIV-1 bnAbs from acute infection through neutralizing antibody development. During their evolution, bnAbs acquire an abundance of improbable amino acid substitutions as a result of nucleotide mutations at variable region sequences rarely targeted by activation-induced cytidine deaminase, the enzyme responsible for antibody mutation. A subset of improbable mutations is essential for broad neutralization activity, and their acquisition represents a key roadblock to bnAb development.

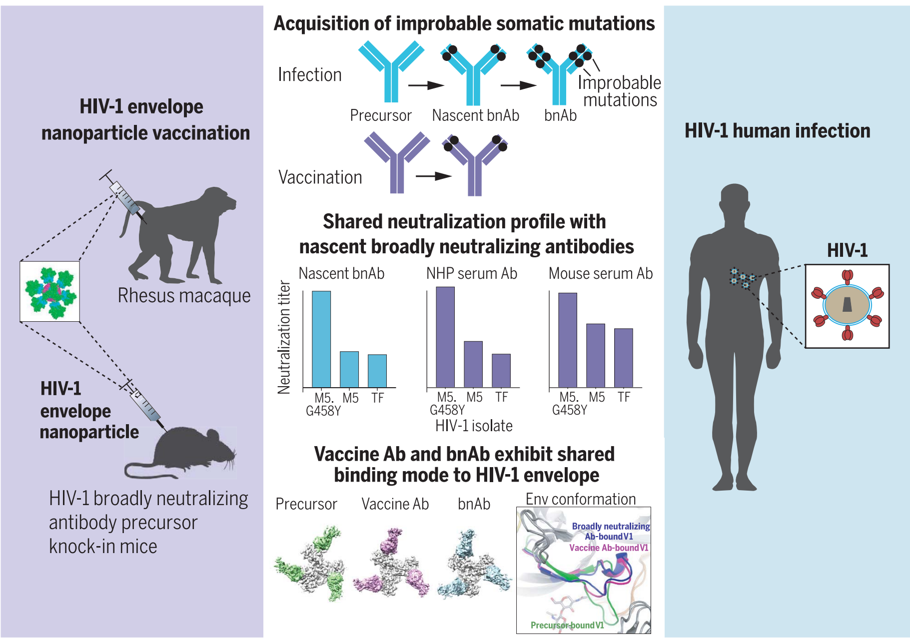
RATIONALE: Current bnAb lineage–based vaccine strategies can initiate bnAb lineage development in animal models but have not specifically elicited the improbable mutations required for neutralization breadth. Induction of bnAbs requires vaccine strategies that specifically engage bnAb precursors and subsequently select for improbable mutations required for broadly neutralizing activity. We hypothesized that vaccination with immunogens that bind with moderate to high affinity to bnAb B cell precursors, and with higher affinity to precursors that have acquired improbable mutations, could initiate bnAb B cell lineages and select for key improbable mutations required for bnAb development.

RESULTS: We elicited serum neutralizing HIV-1 antibodies in human bnAb precursor knock-in mice and wild-type macaques vaccinated with immunogens designed to select for improbable mutations. We designed two HIV-1 envelope immunogens that bound precursor B cells of either a CD4 binding site or V3-glycan bnAb lineage. In vitro, these immunogens bound more strongly to bnAb precursors once the precursor acquired the desired improbable mutations. Vaccination of macaques with the CD4 binding site–targeting immunogen induced CD4 binding site serum neutralizing antibodies. Antibody sequences elicited in human bnAb precursor knock-in mice encoded functional improbable mutations critical for bnAb development. In bnAb precursor knock-in mice, we isolated a vaccine-elicited monoclonal antibody bearing functional improbable mutations that was capable of neutralizing multiple HIV-1 global isolates. Structures of a bnAb precursor, a bnAb, and the vaccine-elicited antibody revealed the precise roles that acquired improbable mutations played in recognizing the HIV-1 envelope. Thus, our immunogens elicited antibody responses in macaques and knock-in mice that exhibited the mutational patterns, structural characteristics, or neutralization profiles of nascent broadly neutralizing antibodies.

CONCLUSION: Our study represents a proof of concept for targeted selection of improbable mutations to guide antibody affinity maturation. Moreover, this study demonstrates a rational strategy for sequential immunogen design to circumvent the difficult roadblocks in HIV-1 bnAb induction by vaccination. We show that immunogens should exhibit differences in affinity across antibody maturation stages where improbable mutations are necessary for the desired antibody function. This strategy of selection of specific antibody nucleotides by immunogen design can be applied to B cell lineages targeting other pathogens where guided affinity maturation is needed for a protective antibody response. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay7199



Overcoming somatic mutation roadblocks to advance broadly neutralizing HIV-1 antibody (bnAb) development. Vaccination of animal models with engineered HIV-1 immunogens generated antibodies that acquired functional improbable mutations critical for virus neutralization. The lack of envelope selection of improbable mutations is a roadblock for bnAb development. Vaccine-elicited antibodies exhibited neutralization activity similar to that of intermediate-stage bnAbs. Structural studies showed a vaccine-elicited neutralizing antibody bound to HIV-1 envelope in a manner similar to that of a mature bnAb.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: kevin.saunders@duke.edu (K.O.S.); barton.haynes@duke.edu (B.F.H.); alt@enders.tch.harvard.edu (F.A.)

Cite this article as K. O. Saunders *et al.*, *Science* **366**, eaay7199 (2019). DOI: 10.1126/science.aay7199

RESEARCH ARTICLE

HIV VACCINES

Targeted selection of HIV-specific antibody mutations by engineering B cell maturation

Kevin O. Saunders[1,2,3]*†, Kevin Wiehe[4]*, Ming Tian[5]*, Priyamvada Acharya[1]*, Todd Bradley[4], S. Munir Alam[4], Eden P. Go[6], Richard Scearce[4], Laura Sutherland[4], Rory Henderson[4], Allen L. Hsu[7], Mario J. Borgnia[7], Haiyan Chen[4], Xiaozhi Lu[4], Nelson R. Wu[4], Brian Watts[4], Chuancang Jiang[4], David Easterhoff[4], Hwei-Ling Cheng[5], Kelly McGovern[5], Peyton Waddicor[5], Aimee Chapdelaine-Williams[5], Amanda Eaton[4], Jinsong Zhang[4]‡, Wes Rountree[4], Laurent Verkoczy[4]‡, Mark Tomai[8], Mark G. Lewis[9], Heather R. Desaire[6], Robert J. Edwards[4], Derek W. Cain[4], Mattia Bonsignori[4], David Montefiori[1], Frederick W. Alt[5]†, Barton F. Haynes[3,4]†

The design of immunogens to direct antibody maturation is a major goal for vaccine development. One roadblock preventing HIV-1 vaccine design is the need for broadly neutralizing antibodies (bnAbs) to acquire somatic mutations rarely made by activation-induced cytidine deaminase (AID). We designed immunogens that bind with higher affinity to antibodies with improbable mutations compared to unmutated precursor antibodies. In knock-in mice, such immunogens engaged unmutated bnAb precursors, selected for functional improbable mutations, and induced neutralizing antibodies. Structural studies revealed how bnAb precursors interact with the envelope protein (Env) and the functions of the elicited improbable mutations. In macaques, the CD4 binding site–targeting immunogen induced potent CD4 binding site–neutralizing antibodies. Our immunogen design strategy may allow for the delineation of sequential immunogens to direct bnAb development for HIV-1.

To date, HIV-1 vaccination has not resulted in the induction of high titers of potent HIV-1 broadly neutralizing antibodies (bnAbs) (*1*, *2*).

[1]Human Vaccine Institute and Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA. [2]Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, NC 27710, USA. [3]Department of Immunology, Duke University School of Medicine, Durham, NC 27710, USA. [4]Human Vaccine Institute and Department of Medicine, Duke University School of Medicine, Durham, NC 27710, USA. [5]Howard Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children's Hospital, Department of Genetics, Harvard Medical School, Boston, MA 02115, USA [6]Department of Chemistry, University of Kansas, Lawrence, KS 66049, USA. [7]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC 27709, USA. [8]Corporate Research Materials Lab, 3M Company, St. Paul, MN 55144, USA. [9]Bioqual, Rockville, MD 20850, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: kevin.saunders@duke.edu (K.O.S.); barton.haynes@duke.edu (B.F.H.); alt@enders.tch.harvard.edu (F.A.) ‡Present address: San Diego Biomedical Research Institute, San Diego, CA 92121, USA.

bnAbs are disfavored by immune tolerance mechanisms because of their unusually long complementarity-determining regions (CDRs), autoreactivity, and polyreactivity (*3*, *4*). In addition, bnAbs have high frequencies of somatic mutation resulting from extended rounds of affinity maturation (*5*–*7*). Antibody somatic mutation is mediated by activation-induced cytidine deaminase (AID), the enzyme that deaminates cytidine to uridine and can lead to nucleotide substitution during DNA repair (*8*). As a result of the preferential targeting of AID to specific sequence motifs, mutability varies greatly among positions within an antibody sequence (*9*). We recently used the computational program Antigen Receptor Mutation Analyzer for Detection of Low-likelihood Occurrences (ARMADiLLO) to determine that bnAbs are enriched for somatic mutations that occur at variable region sequences not routinely targeted by AID or that require multiple changes to the germline codon (*10*). A subset of these improbable mutations are required for broad neutralization activity and therefore represent key roadblocks for the development of bnAbs (*10*–*12*). These obstacles have led to the hypothesis that vaccine strategies are needed that direct the immune system to expand B cell lineages that are usually rare and select for disfavored antibody traits (*4*). At the foundation of this vaccine strategy is the specific engagement of germline precursors of neutralizing antibodies and the selection of improbable mutations by strong antigenic selection (*10*, *11*, *13*). Recent studies have reported variable results in induction of V3-glycan bnAb B cell lineage precursors, either starting from mutated precursors (*14*) or induction of precursors in macaques (*15*); however, to date, there are no reports of vaccination specifically eliciting improbable mutations above the frequency with which they would be expected to occur in the absence of selection (*10*, *11*). Learning the rules to precisely select for specific antibody somatic mutations would improve the design of vaccines or therapeutics for infectious diseases, autoimmunity, inflammatory diseases, and malignant diseases (*16*, *17*).

One of the major bnAb epitopes on HIV-1 Env is the V3-glycan site, which consists of the base of the third variable (V3) loop and surrounding glycans (*18*–*21*). The DH270 bnAb B cell lineage, isolated from an HIV-infected individual, targets the V3-glycan epitope on HIV-1 Env (*18*). We computationally inferred the clonal history of the DH270 lineage, including the naïve B cell V(D)J rearrangement of the DH270 lineage termed the unmutated common ancestor (UCA) (*18*). The computational reconstruction of the DH270 lineage showed that the lineage initially progressed from the UCA precursor antibody to an early intermediate antibody (IA) designated DH270 IA4 (*18*). This initial affinity maturation step for the DH270 lineage included the acquisition of four amino acid changes [$Gly^{31} \rightarrow$ Asp, $Met^{34} \rightarrow$ Ile, $Ser^{55} \rightarrow$ Thr, and $Gly^{57} \rightarrow$ Arg (G57R)] and one amino acid change [$Ser^{27} \rightarrow$ Tyr (S27Y)] in the heavy and light chain variable regions, respectively (*18*). The DH270 IA4 antibody bearing these mutations had the ability to neutralize a subset of HIV-1 isolates representative of the globally circulating virus population (heterologous viruses) (*18*). Two of the five amino acid changes in DH270 IA4 were the result of mutations at AID hot spots, and their acquisition did not contribute to heterologous HIV-1 neutralization (*18*). In contrast, the G57R amino acid change alone is sufficient for heterologous HIV-1 neutralization and is the result of a nucleotide mutation within a disfavored AID cold spot, making its acquisition highly improbable (*10*, *18*). Thus, immunogens are needed that can specifically select for the G57R amino acid change in order for vaccines to elicit DH270-like V3-glycan antibody responses.

The CD4 binding site bnAb lineage CH235 was isolated from the CH505 HIV-1–infected individual (*22*). Unlike most potent CD4 binding site bnAbs that mimic CD4, the CH235 lineage does not require difficult-to-elicit insertions or deletions for broad neutralization (*22*, *23*). Instead, the most potent bnAbs in the CH235 lineage had many improbable mutations, which suggested that the primary obstacle in CH235 lineage development is accumulation of improbable mutations (*10*). When the CH235 clonal lineage was reconstructed, a key early improbable functional mutation, $Lys^{19} \rightarrow$ Thr (K19T) in the heavy chain variable region, was identified. The K19T amino acid change improved binding to Env by the CH235 bnAb precursor and was necessary for neutralization of HIV-1 by CH235 bnAbs (*10*). The generation of an immunogen capable of eliciting the K19T change would overcome the first roadblock in the development of the CH235 lineage, and it could be applicable to other antibodies in the CH235 (also called

8ANC131) bnAb class because they all encode K19T changes as well.

Here, we designed immunogens that select for specific functional improbable nucleotide mutations required for bnAb development, thus demonstrating proof-of-concept for directing DH270 V3-glycan and CH235 bnAb B cell lineage affinity maturation. These immunogens for two different bnAb lineages can overcome the two initial roadblocks—engagement of the bnAb precursor and selection of functional improbable mutations—and are key components of a sequential HIV-1 vaccine regimen.

Engagement of V3-glycan bnAb precursor B cell receptors

We previously determined that recombinant HIV-1 Env gp120s produced from sequences isolated from the HIV-1–infected individual who developed the DH270 bnAb lineage were incapable of binding to the UCA (*18*). Computational modeling of the structure of the bnAb DH270.6, the most potent DH270 bnAb lineage member, when bound to trimeric Env showed steric hindrance by an N-linked glycan present at amino acid 137 in the first variable region (V1) of HIV-1 Env (fig. S1A). Using an autologous Env from CH848, we removed glycosylation sites near position 137—the Asn^{133} and Asn^{138} (N133 and N138) glycans. The DH270 UCA neutralized the CH848.D949.10.17 virus with the N133 and N138 glycans removed (fig. S1C; 50% inhibitory concentration IC_{50} = 1.7 μg/ml), but the same virus with the glycans present was not neutralized (Fig. 1A). Removing the V1 glycans from JR-FL was not sufficient for DH270 UCA neutralization sensitivity (fig. S1B). This difference in neutralization sensitivity between JR-FL lacking V1 glycans and CH848 10.17 lacking V1 glycans indicated that V1 glycan removal alone was not sufficient for making all viruses sensitive to DH270 UCA neutralization (fig. S1B).

For immunizations, soluble, well-folded CH848 gp140 SOSIP Env trimers with (10.17) and without (10.17DT) the N133 and N138 glycans present were produced (fig. S1C and figs. S2 to S4). We initially generated stabilized SOSIPs using three different published strategies (*24*–*26*) but found them to have very similar trimer conformation and antigenicity (fig. S2). We selected the SOSIPv4.1 version with the lowest non-neutralizing antibody binding for further study. The CH848 10.17DT Env trimer bound to the DH270 UCA antigen-binding fragment (Fab) with an equilibrium binding constant (K_D) of 532 nM (Fig. 1B). The presence of V1 glycans reduced the binding affinity by a factor of 5 between the DH270 UCA and CH848 10.17DT–stabilized SOSIP Env (fig. S1D). The K_D of CH848 10.17DT improved to 118 nM for the DH270 IA4 Fab, although IA4 differed from the UCA by only five amino acids (Fig. 1B). The difference in binding affinity suggested that 10.17DT could select these five amino acid changes. The selection of these amino acids would advance antibody affinity maturation toward neutralization breadth. When the DH270 UCA immunoglobulin M (IgM) was expressed as an IgM B cell receptor (BCR) on the surface of a Ramos B cell line, 10.17DT Env trimers bound with sufficient avidity to cross-link the receptor and induce calcium flux (Fig. 1C). Removal of the N133 and N138 V1 glycans resulted in an increase in the percentage of high-mannose glycans at Asn^{332} on the CH848 10.17DT Env trimer (fig. S3). Because DH270-lineage antibodies bind the Asn^{332} glycan and are high mannose–reactive, the increase in high mannose likely contributed to the improved binding to the DH270 UCA (figs. S1D and S3). Therefore, V1 glycans were inhibitory for DH270 antibody binding to envelope trimers, and their removal generated an envelope capable of interacting with soluble and membrane-bound DH270 UCA.

Structural characterization of DH270 UCA bound to CH848 10.17DT Env

To provide atomic-level information on the interaction between DH270 UCA and the CH848 10.17DT Env, we determined the structure of the antigen-binding fragment (Fab) of DH270 UCA in complex with CH848 10.17DT SOSIP by cryo–electron microscopy (cryo-EM) to an overall resolution of 4.2 Å (Fig. 1, D to H, figs. S5 and S6, and table S1). In cryo-EM reconstructions, we observed that the Env trimer was engaged by three Fabs in a 1:1 Env protomer/Fab ratio (Fig. 1D). The DH270 UCA Fab bound a proteoglycan epitope composed of variable loops 1 (V1) and 3 (V3) and surrounding glycans at positions 301 and 332 (Fig. 1, E to H). Of the ~2212 $Å^2$ total surface area buried at the Env-antibody interface, ~57% was contributed by interactions of the antibody with Env glycans. Whereas glycan 301 contacted only the light chain (total buried interface area ~250 $Å^2$), glycan 332 made contacts with both the heavy and light chains, burying a total of 1047 $Å^2$ at its interface, with about two-thirds of the interactive surface contributed by the heavy chain (Fig. 1F).

DH270 UCA contacted the V3 loop by means of its HCDR1, HCDR2, and HCDR3 loops. The interaction was focused on the conserved V3 GDIK/R (Gly-Asp-Ile-Lys/Arg) motif, with the side chain of HCDR1 Tyr^{33} in the fitted coordinates placed favorably to engage in a potential hydrogen bond with the side chain of the conserved Asp^{325} of the GDIK sequence (Fig. 1G). The N133, N138–deleted V1 loop was stacked against the V3 loop and adopted a β hairpin–like structure with the tip of the hairpin stabilized by interactions with the HCDR2 loop. In contrast to the N156 glycan (Fig. 1E), the presence of a glycan at position 138 in the V1 loop would have created steric clashes with the HCDR2 loop of the bound DH270 UCA (Fig. 1, E, G, and H). The N133D mutation was distal from the binding site of DH270 UCA but may modulate the conformation of the V1 loop. In summary, the structure of the Env-DH270 UCA complex provided an atomic-level understanding of the interaction of the DH270 UCA with the Env CH848 10.17DT by revealing key contacts with conserved elements of the V3 loop and the surrounding glycans, and by showing how the engineered V1 loop engaged the DH270 UCA.

Generation and characterization of DH270 UCA knock-in mice

To evaluate the ability of the CH848 10.17DT Env to bind to the DH270 UCA BCR on primary B cells and to test CH848 10.17DT Env immunogenicity, we developed a humanized mouse model with the V_HDJ_H and $V_\lambda J_\lambda$ regions of the DH270 UCA knocked in (fig. S7A). As shown in fig. S7B, the DH270 UCA V_H + V_L heterozygous knock-in (KI) mice had reduced numbers of B cells relative to T cells, which served as an internal reference. In control mice without KI genes, the ratio of B cells to T cells was 1.4; this was reduced to about 0.4 in DH270 UCA heterozygous KI mice. An obvious increase in IgM^+ IgD^{lo} immature B cell populations was observed in the spleen of DH270 UCA KI mice relative to control mice (figs. S7C and S8C), suggesting a slowdown or block in B cell maturation. To determine the proportion of B cells expressing DH270 UCA V_H, we examined allotypic markers of the two IgH alleles: DH270 UCA V_H was expressed as IgM^a, whereas mouse heavy chains were expressed as IgM^b. As shown in fig. S7C, about 30% of splenic B cells expressed IgM^a and more than 50% of B cells expressed IgM^b. In theory, expression of a pre-rearranged V_H exon should preclude the rearrangement of the other IgH^b allele (*27*). However, KI heavy chains of autoreactive antibodies can be deleted via V_H replacement (*28*–*31*), thereby freeing the other IgH allele for expression. DH270 UCA V_L was expressed as a κ light chain, but the KI mice expressed 2.5 times as many Ig λ^+ B cells as normal mice, which is indicative of receptor editing (*32*–*34*). Overall, the B cell phenotype of the DH270 UCA KI mice is consistent with the concept that the expression of some bnAbs is under negative selection by tolerance control mechanisms (*35*). However, the extent of B cell deletion in the DH270 UCA model is less severe than in some previously reported bnAb KI mouse models (*36*–*38*). The DH270 UCA BCR was functional, as the 10.17DT-stabilized SOSIP trimer induced calcium flux in splenic B cells from each of the three homozygous DH270 UCA ($V_H^{+/+}$, $V_L^{+/+}$) KI mice tested (fig. S9). Calcium flux was not detectable when the 10.17-stabilized SOSIP was used as the antigen (fig. S9). Because tolerance

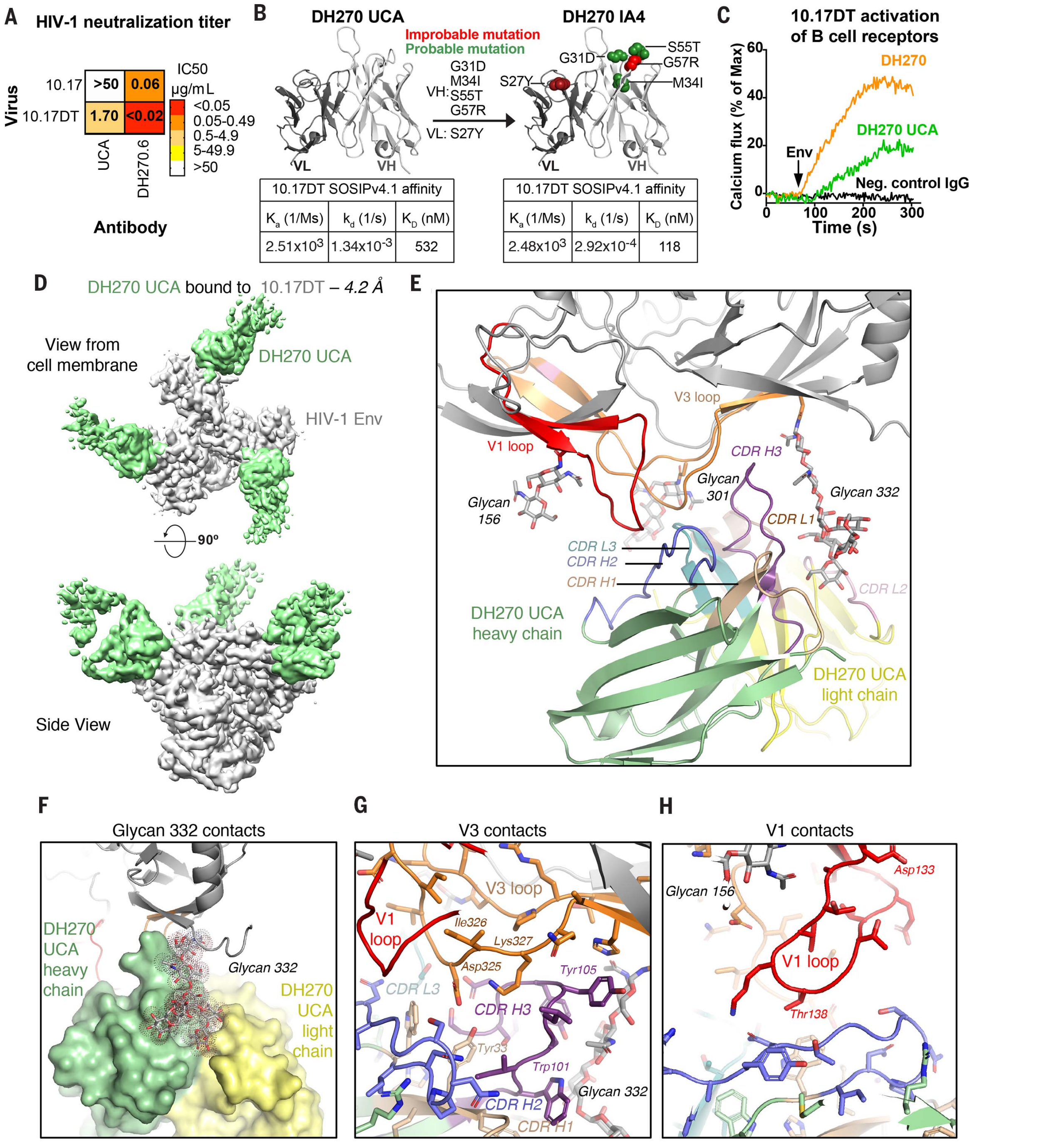


Fig. 1. Structural characterization and development of an HIV-1 Env antigen capable of binding to the DH270 unmutated common ancestor (UCA). (**A**) Antibody neutralization of TZM-bl cell infection was compared for HIV-1 pseudoviruses encoding CH848 10.17 gp160 Env or the same Env with glycosylation sites at Asn[133] and Asn[138] (N133 and N138) removed by mutating them to Asp[133] and Thr[138] (10.17DT). Neutralization titers are shown as IC_{50} values. (**B**) The binding affinities of DH270 UCA and the first intermediate antibody in the DH270 lineage (IA4) for CH848 10.17DT–stabilized Env trimers was determined by surface plasmon resonance. The somatic mutations acquired in the heavy chain variable region (V_H) and light chain variable region (V_L) of DH270 IA4 are shown above and below the arrow, respectively. Improbable and probable somatic mutations were modeled on the DH270 UCA Fab crystal structure (PDB ID 5U0R) as red and green residues, respectively. (**C**) CH848 10.17DT calcium flux induction in DH270 UCA (green), DH270.1 (orange), and negative control (black) IgG-expressing Ramos B cells. (**D**) Cryo-EM reconstruction (4.2 Å resolution) of DH270 UCA (green) bound to CH848 10.17DT–stabilized Env trimers (gray), segmented by components. (**E**) Zoomed-in view of the interactive region between DH270 UCA and CH848 10.17DT. HIV-1 Env gp120 is shown in gray, with the V3 loop in orange and the V1 loop in red. Glycans are shown in stick representation and colored by elements, with oxygen atoms in red and nitrogen atoms in blue. The heavy chain of DH270 UCA is shown in green and light chain in yellow. Color code for complementarity-determining regions (CDRs) of the antibody: wheat, CDR H1; blue, CDR H2; purple, CDR H3; brown, CDR L1; pink, CDR L2; teal, CDR L3. DH270 UCA contacted the V1 loop, the V3 loop, and the surrounding glycans, with contacts made by both the heavy and light chains of the antibody. (**F** to **H**) Zoomed-in views of the structure showing details of the contacts DH270 UCA makes with glycan 332, shown in stick representation overlaid with dots and colored by element, with the DH270 UCA shown in surface representation (F), the V3 loop (G), and the V1 loop (H).

control mechanisms may also curtail DH270 precursor expression in humans, this DH270 V_H + V_L heterozygous mouse provides a relevant and challenging model to test our immunogens.

Nanoparticle immunization of DH270 UCA KI mice

Multimeric immunogens have been reported to be superior to soluble monomers for inducing humoral responses (15, 39–46). Therefore, we sought to generate a nanoparticle immunogen arraying the 10.17DT SOSIPv4.1 version of the Env trimer. The 10.17DT Env was chosen over the wild-type 10.17 on the basis of 10.17DT's stronger binding affinity (by a factor of 5) for the DH270 UCA (fig. S1D), detectable induction of calcium flux (fig. S9), and more frequent elicitation of heterologous-neutralizing serum antibodies in DH270 UCA KI mice (fig. S10). To ensure that well-folded, native-like Env trimers were present on a multimeric Env trimer nanoparticle, we purified well-folded trimers and used sortase A to site-specifically ligate them to ferritin nanoparticles (Fig. 2A and fig. S11). HIV-1 Env trimers on the nanoparticle had the same bnAb-specific antigenic profile as the soluble Env trimer, in that neutralizing antibodies but not non-neutralizing antibodies bound to the trimer (Fig. 2B and fig. S11). The CH848 10.17DT Env trimer nanoparticles cross-linked the BCR on mature and immature splenic B cells from homozygous DH270 UCA ($V_H^{+/+}$, $V_L^{+/+}$) KI mice and induced intracellular calcium flux (Fig. 2C).

Next, we immunized heterozygous DH270 UCA ($V_H^{+/-}$, $V_L^{+/-}$) KI mice with CH848 10.17DT Env trimer nanoparticles. In this mouse model, about 12% of splenic follicular B cells bind CH848 10.17DT, making this model more physiologic than previous KI model systems where most B cells express the KI variable regions (14) (Fig. 2D and fig. S8C). Mice that received the Env-trimer nanoparticle had significantly higher numbers of germinal center B cells and follicular helper T cells relative to adjuvant only–immunized mice (Fig. 2E). The Env-trimer nanoparticle induced N332-directed antibodies and tier 2 autologous neutralizing antibodies against CH848 10.17DT and CH848 10.17 viruses after two or three immunizations, respectively (Fig. 2F and fig. S12). Heterologous neutralizing antibodies were also detectable in four of five CH848 10.17DT Env-immunized mice, but not in those animals that received adjuvant only (fig. S12). Thus, in an in vivo setting where DH270 UCA precursors represented a minority of B cells, high-quality Env nanoparticles induced serum V3-glycan antibodies.

Selection of improbable mutations by CH848 10.17DT immunization

We next determined whether 10.17DT vaccination could select for improbable somatic mutations that occurred during the natural evolution of DH270 in the HIV-1 infected individual. 10.17DT-stabilized envelope trimer bound to the DH270 UCA with the improbable G57R mutation with affinity improvement by a factor of 4 relative to the DH270 UCA (K_D = 132 nM and 558 nM, respectively; figs. S13, A and B, and S14, A to C). We hypothesized that this improvement in binding affinity would provide the necessary affinity difference to select for the G57R mutation. Indeed, next-generation sequencing of heavy chain variable regions (V_H NGS) from the vaccinated mice demonstrated that the G57R mutation occurred more frequently in CH848 10.17DT nanoparticle-immunized DH270 UCA KI mice than in adjuvant only–treated mice (Fig. 2G, top). Immunization with the CH848 10.17DT nanoparticle increased the median frequency of the G57R mutation within DH270 sequences by a factor of 200 after nanoparticle immunization (Fig. 2G, bottom).

In addition to the G57R improbable mutation, DH270 IA4 encodes a light chain variable region (V_L) S27Y improbable mutation (Fig. 1B and fig. S13C). Addition of the S27Y mutation to the DH270 UCA enabled accommodation of V1 glycan, as manifested by more potent neutralization of CH848 10.17 virus with the V1 glycans present (fig. S13A). We isolated single memory B cells that bound to the CH848 10.17 Env trimer from Env nanoparticle–vaccinated mice (fig. S15). Of the 10.17 Env trimer-reactive antibodies, 66% were derived from the DH270 antibody B cell lineage and 99% of these DH270 antibodies were somatically mutated (Fig. 3, A and B, and fig. S16). The somatic mutations included improbable and probable mutations (figs. S16 to S18). Among the DH270 antibodies from the nanoparticle-immunized mice, antibody DH270.mu1 encoded both V_H G57R and V_L S27Y mutations (Fig. 3D and fig. S18). The DH270.mu1 antibody neutralized autologous tier 2 CH848 viruses as well as heterologous HIV-1 6101, Q23, and 92RW020 viruses with titers nearly identical to antibody DH270 IA4 (Fig. 3E). DH270.mu1 neutralized 16 of 23 heterologous isolates tested, including heterologous tier 1B and tier 2 viruses (Fig. 3F). Hence, vaccination induced a heterologous tier 2 neutralizing DH270 antibody that possessed the G57R and S27Y improbable mutations.

Accumulation of additional bnAb mutations in vaccine-elicited antibodies

In addition to S27Y and G57R, the V_H and V_L amino acid sequences of the 134 nanoparticle-induced DH270 antibodies included other somatic mutations present in DH270.1 (figs. S16 and S18). More specifically, 16 of 19 somatic amino acid changes present in DH270.1 were also present among the vaccine-induced antibodies (Fig. 3G). However, no single vaccine-induced antibody had acquired all 16 observed V_H mutations in combination; instead, they were distributed among the 134 antibodies. Although 90% of the vaccine-induced antibodies shared two somatic mutations with DH270.1, the highest number of shared mutations accumulated on a single vaccine-induced DH270 V_H sequence was six; this occurred in 6% of the observed monoclonal DH270 antibody V_H sequences (Fig. 3H). Similar to the V_H sequences, 6 of 10 amino acid changes present in DH270.1 light chain were elicited by 10.17DT nanoparticles. Of the three V_H mutations that were not sampled, none were improbable, and of the four V_L mutations that were not sampled, two were improbable (figs. S17 and S19).

To examine more deeply the accumulation of low V_H somatic mutations, we performed V_H next-generation sequencing on splenic B cells. We examined DH270 sequences for the co-occurrence of R98T and G57R because R98T has also been shown to be critical for DH270 interaction with glycan (47). The co-elicitation of both R98T and G57R was observed for all mice immunized with CH848 10.17DT nanoparticle at a higher frequency than adjuvant alone (fig. S20). The highest percentage of DH270 sequences with both mutations from any one mouse was 3.7% of DH270 sequences (fig. S20). We also analyzed DH270 sequences from the mice for the occurrence of all four DH270 IA4 V_H mutations in single antibodies. All mice immunized with CH848 10.17DT nanoparticle generated DH270 sequences containing all four somatic mutations present in DH270 IA4, but administration of adjuvant alone did not generate such sequences (fig. S20). Thus, the CH848 10.17DT nanoparticle selected for limited combinations of mutations present in the initial stages of the DH270 lineage maturation, but again did not elicit antibodies with all 19 somatic mutations observed in DH270.1. These data indicate the importance of sequential boosts to select for the improbable mutations required for full bnAb capacity.

We hypothesized that without strong antigenic selection, acquiring all improbable mutations observed in DH270.1 (18) in one antibody would be a highly unlikely event. Under a scenario in which only four improbable mutations are required (fig. S19, A and B), and given the assumption of a constant rate of one improbable mutation per 12 weeks (the duration of immunization in nanoparticle-immunized mice), it would take 120 weeks of biweekly immunization of mice to reliably acquire four improbable mutations (fig. S19C), or multiple decades of immunization in humans given the age relation between mice and humans (48). Although B cell evolution does not proceed under a constant mutation rate (7, 49), these data suggest that

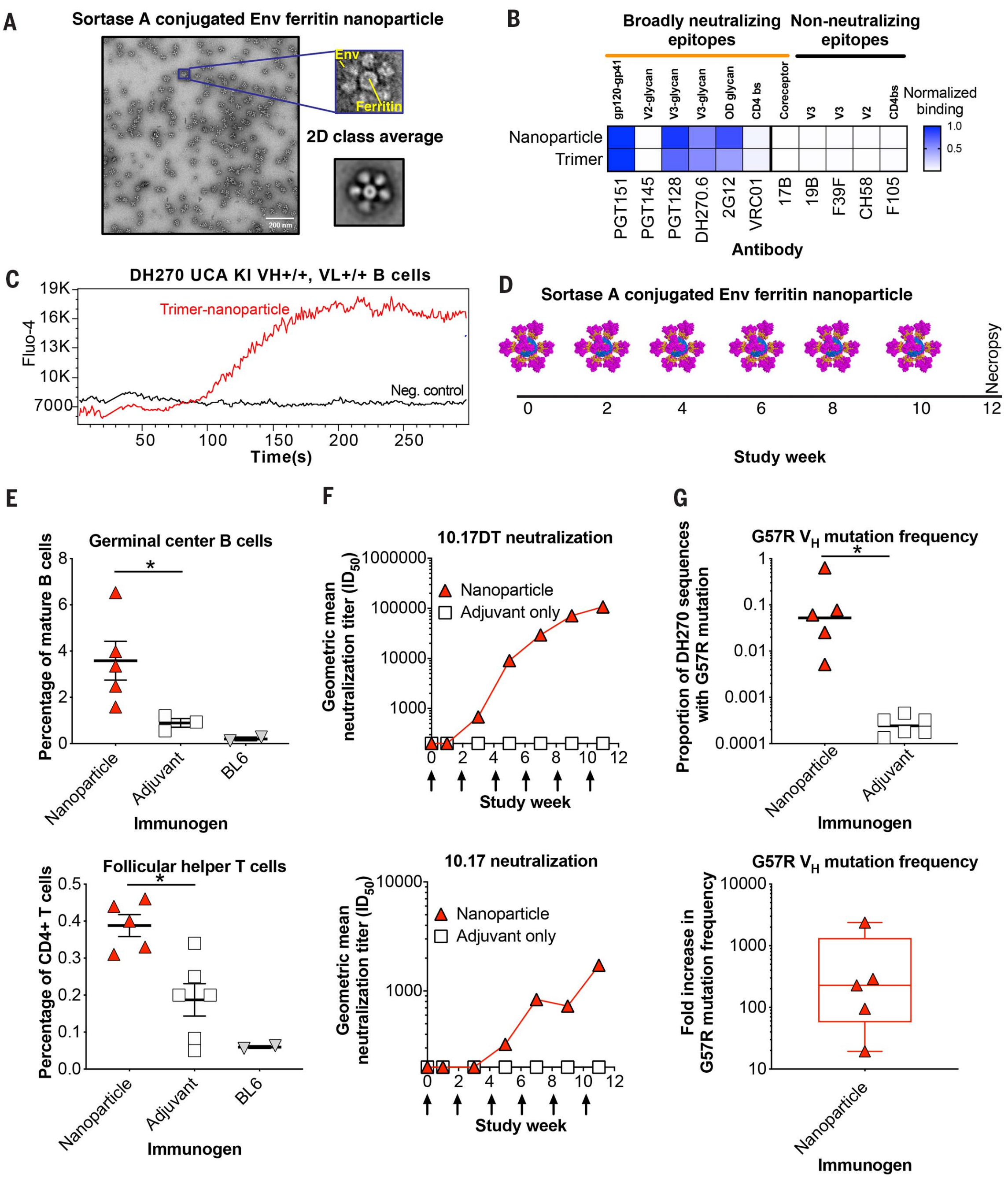


Fig. 2. CH848 10.17DT nanoparticle immunization of DH270 UCA knock-in mice induces serum neutralizing antibody responses and the improbable G57R mutation. (**A**) Negative-stain electron microscopy of CH848 10.17DT–stabilized Env trimers conjugated to ferritin nanoparticles. A magnified image of the Env nanoparticles (top) and 2D class average (bottom) of the Env nanoparticles are shown at the right. (**B**) Antigenic profile of CH848 10.17DT–stabilized Env trimers conjugated to ferritin nanoparticles or unconjugated. Binding was determined by biolayer interferometry and normalized to PGT151 binding values. The normalized binding is shown in the heat map, with a value of 1 being equal to PGT151 binding. OD glycan, outer domain glycans in the high-mannose patch. (**C**) CH848 10.17DT Env trimer nanoparticle (red) induction of calcium flux in homozygous ($V_H^{+/+}$, $V_L^{+/+}$) DH270 UCA double knock-in (double KI) mouse B cells. (**D**) Heterozygous ($V_H^{+/-}$, $V_L^{+/-}$) DH270 UCA double KI mouse immunization regimen with CH848 10.17DT SOSIP trimer nanoparticles. (**E**) Quantification of the frequency of splenic germinal center B cells (top) and follicular helper T cells (bottom) in vaccinated DH270 UCA mice. Unvaccinated, wild-type C57BL/6 mice (BL6) served as baseline controls. Means ± SEM are shown by black bars. **P* < 0.05 (Wilcoxon exact test). (**F**) Serum antibody neutralization of autologous tier 2 CH848 10.17DT (top) and CH848 10.17 (bottom) virus infection of TZM-bl cells. Group geometric mean ID_{50} titers are shown (nanoparticle, *n* = 5; adjuvant only, *n* = 6). Arrows indicate immunization time points. (**G**) Top: Enumeration by V_H next-generation sequencing of the frequency of unique DH270 sequences encoding the G57R amino acid change. Frequencies of G57R were determined in total splenocytes 1 week after the sixth immunization. Each mouse is shown by an individual symbol; horizontal bars indicate the group mean. Bottom: Relative (fold) increase in G57R frequency in heterozygous DH270 UCA KI mice immunized with CH848 10.17DT Env nanoparticle and adjuvant compared to adjuvant-only control mice. Box-and-whisker plots indicate the minimum, maximum, median, and interquartile range fold increase for each group. Each symbol represents an individual mouse. **P* < 0.05 (Wilcoxon exact test).

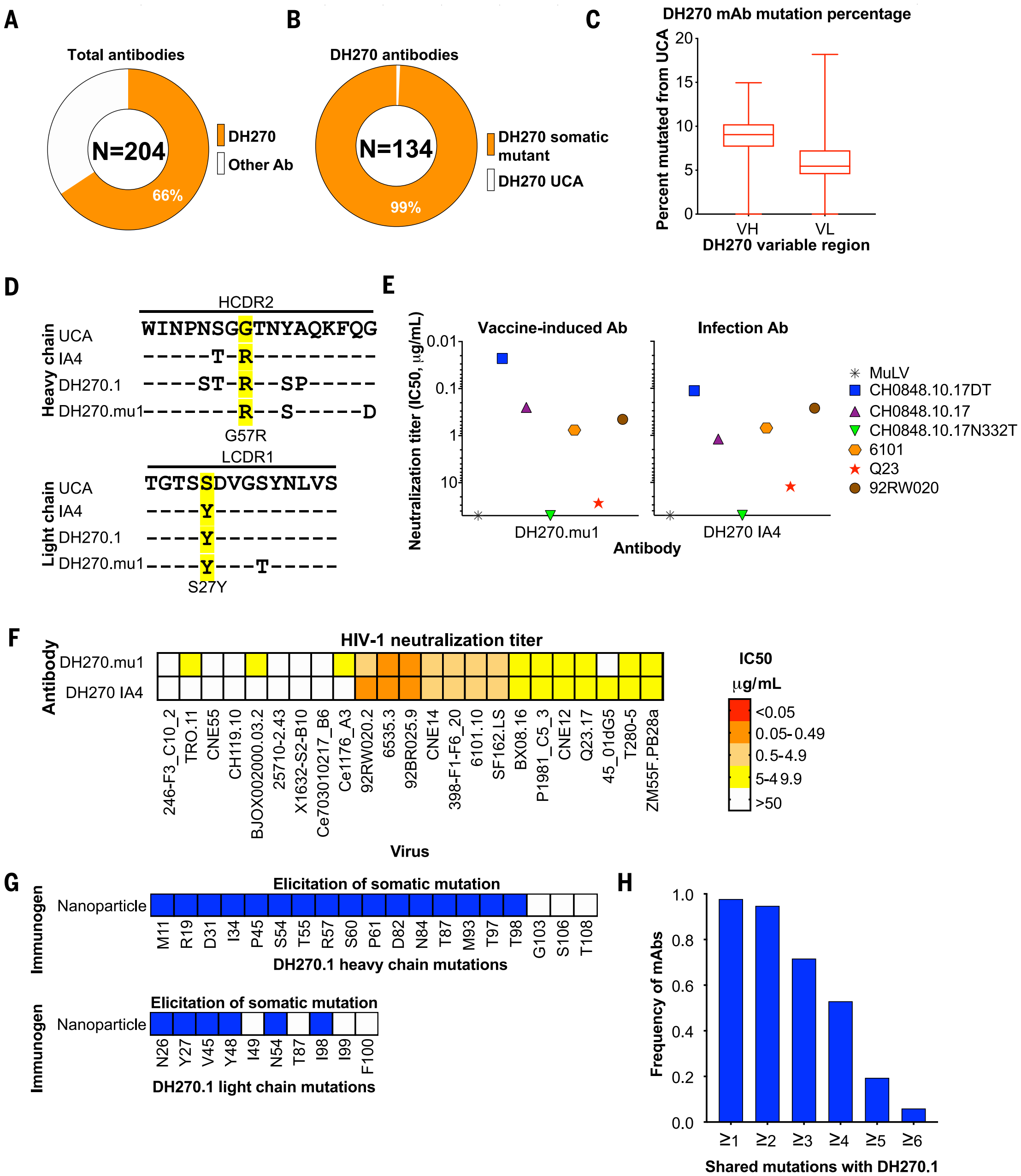


Fig. 3. CH848 10.17DT Env trimer nanoparticle immunization elicits neutralizing antibody DH270.mu1 that encodes both S27Y and G57R improbable amino acid changes. (**A**) The percentage of CH848 10.17–specific B cells sorted from CH848 10.17DT Env trimer nanoparticle–immunized mice that encode DH270 V_H and V_L regions. CH848 10.17–specific B cells encoding one or more mouse Ig chains were classified as other antibodies. Total number of antibodies isolated is shown in the center of the pie chart. (**B**) The percentage of DH270-expressing B cells that have acquired one or more somatic mutations. Total number of DH270 antibodies isolated [based on data in (A)] is shown in the center of the pie chart. (**C**) Amino acid mutation frequency of the DH270 V_H and V_L regions cloned from the sorted CH848 10.17–specific B cells shown in (B). Box-and-whisker plots show the range, median, and interquartile range. (**D**) Env nanoparticle vaccine–induced monoclonal antibody DH270.mu1 possessed both the G57R and S27Y amino acid changes in its HCDR2 and LCDR1 loops, respectively. Shown is an amino acid alignment of select DH270 clonal lineage antibodies and DH270.mu1. The two critical improbable mutations are highlighted in yellow. (**E**) DH270.mu1 neutralizes autologous and heterologous viruses with potency nearly identical to that of DH270 IA4. Neutralization titers are shown as IC_{50} for each individual virus. Left, vaccine-induced antibody; right, DH270 intermediate antibody inferred from the clonal lineage elicited during chronic HIV-1 infections. (**F**) DH270.mu1 possesses neutralization breadth (16/23 or 70% of viruses) similar to that of DH270 IA4. Neutralization IC_{50} titers are shown in a heat map. (**G**) The somatic mutations shared between the vaccine-elicited antibodies and the bnAb DH270.1. The DH270.1 somatic mutations observed after vaccination with 10.17DT trimer nanoparticle are indicated with blue squares; unobserved mutations are shown by white squares. (**H**) The frequency of the total (n = 134) mouse mAbs isolated from mice immunized with the 10.17DT SOSIP nanoparticle that had the indicated numbers of shared mutations with DH270.1. Amino acid abbreviations: A, Ala; D, Asp; F, Phe; G, Gly; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

boosting immunogens designed to specifically select for combinations of key improbable mutations will be necessary to accelerate neutralizing antibody lineage maturation to broad neutralization.

To determine whether the V1 glycan deletions were important for optimal selection of the G57R mutation, we compared elicitation of G57R mutations in mice immunized with CH848 10.17 versus CH848 10.17DT in DH270 UCA KI mice. We found that the 10.17DT SOSIP trimer more consistently elicited higher frequencies of the G57R improbable mutation than did the CH848 10.17 trimer (fig. S14D). Additionally, to determine whether the selection of the G57R mutation by CH848 10.17DT Env was a result specific to this immunogen, we analyzed previously published antibody sequences from KI mice that encode the same human V_H1-2*02 gene as DH270 but were immunized with a different HIV-1 Env immunogen (*50*, *51*). Analysis of next-generation sequencing of V_H regions from V_H1-2*02 KI mice immunized with eODGT8 nanoparticles showed that this immunogen did not elicit the G57R mutation (*50*, *51*). Thus, the selection of the desired G57R mutation is not merely a product of activation of V_H1-2*02–bearing B cells and is therefore the result of immunization of the DH270 UCA KI mice with the CH848 10.17DT Env.

Structural comparison of DH270.mu1 and DH270.6

To provide information on the interaction between the nanoparticle-induced antibody DH270.mu1 (containing two key improbable mutations, G57R and S27Y) and the 10.17DT Env, we determined the structure of the DH270.mu1 Fab in complex with 10.17DT SOSIP trimer by cryo-EM to an overall resolution of 3.5 Å. We also determined the structure of the affinity-matured V3-glycan bnAb DH270.6 (*18*) in complex with CH848 10.17DT SOSIP to an overall resolution of 4.3 Å (Fig. 4, A and B, figs. S21 and S22, and table S1). By comparing the DH270.mu1 complex to the DH270 UCA and the matured DH270.6 bnAb complexes, we were able to assess the structural evolution of the vaccine-induced DH270.mu1 antibody. DH270 UCA, DH270.mu1, and DH270.6 all bound at a 1:1 Env protomer/Fab ratio with similar overall orientations (Fig. 4, A to D, and figs. S6, S21, and S22). All three antibodies used the 20–amino acid HCDR3 to contact the base of the V3 loop by reaching past the N332 glycan (Fig. 1E and Fig. 4, C and D). The most striking difference was observed in the conformation of the V1 loop (Fig. 1E and Fig. 4, C to H). In the DH270 UCA-bound complex, the tip of the short V1 loop was oriented toward the antibody paratope with the HCDR2 loop making van der Waals contacts with the V1 tip (Fig. 1, E, G, and H, and Fig. 4E). In contrast, the V1 loop in the DH270.6-bound complex was positioned roughly perpendicular to the antibody paratope, with the tip of the V1 loop separated by about 14 Å between the DH270 UCA-bound and DH270.6-bound conformations (Fig. 4, E to H). The V1 loop in the DH270.mu1-bound complex adopted a conformation similar to that in the DH270.6-bound complex, and distinct from the V1 loop conformation observed in the DH270 UCA–bound complex (Fig. 4, E to H). The altered disposition of the V1 loop, and its movement away from the HCDR2 loop in the DH270.6- and DH270.mu1-bound complexes, allowed access to the conserved GDIK motif in the V3 loop by the improbable mutation G57R, via hydrogen bonds between the side chain of Arg^{57} of the HCDR2 loop and the main-chain carbonyl of residue Gly^{324} in the GDI(R/K) motif of the V3 loop (Fig. 4, G and H, and figs. S21B and S22B). We also examined the LCDR1 loop, which is the site of the improbable S27Y tyrosine mutation that was acquired in both the mature DH270.6 bnAb and the elicited DH270.mu1 antibody. The tyrosine substitution facilitated an interaction with the glycan 301 (Fig. 4, I to K). Thus, the three cryo-EM structures revealed that the V1 loop lacking N133 and N138 glycans was conformationally flexible, adopted two distinct conformations when bound to the DH270 UCA versus the DH270.6 mature bnAb, and, when bound to the elicited DH270.mu1 antibody, resembled the V1 loop conformation in the DH270.6-bound complex.

Induction of precursors of broad and potent CD4 binding site CH235-class neutralizing antibodies with improbable mutations

To generalize the selection of improbable mutations to additional bnAb B cell lineages, we designed immunogens for the CD4bs bnAb lineage CH235 that was derived from the CH505 HIV-1–infected individual (*22*). The CH235-lineage bnAbs have many improbable mutations, including a functional improbable mutation, K19T in V_H (*10*). The K19T mutation is sufficient on its own to expand neutralization breadth against autologous CH505 viruses by the CH235 UCA (Fig. 5A). However, only low-affinity autologous envelopes have been shown to bind to the CH235 UCA (fig. S25A) (*23*). In recent work, LaBranche and colleagues showed that the CH505 M5 envelope with a G458Y (M5.G458Y) mutation was bound by the CH235 UCA with high affinity (*52*). The apparent affinity was improved by a factor of 9 by generating the envelope in $GnTI^{-/-}$ cells, which restricted glycan processing beyond $Man_5GlcNAc_2$ (Man, mannose; GlcNAc, *N*-acetylglucosamine) (*52*). M5.G458Y-stabilized SOSIP gp140 bound with an apparent affinity of 9 nM for the CH235 UCA and 0.1 nM for the CH235 early bnAb that included the K19T change (*52*). Thus, we identified a high-affinity Env immunogen that had the potential to select for the critical K19T change.

Generation and characterization of CH235 UCA knock-in mice

To evaluate immunogen selection of the K19T mutation, we generated CH235 UCA heterozygous ($V_H^{+/-}$, $V_L^{+/-}$) KI mice in the same way as the DH270 UCA model. In this mouse line, the ratio of splenic B cells versus T cells was close to normal (fig. S23A). The IgM/IgD profile of splenic B cells was comparable to that in control mice, with elevation of the IgM^+IgD^{lo} population by less than a factor of 2 (figs. S23B and S24C). There was no detectable population of surface Ig^{lo} B cells, which are usually anergic. About 30% of B cells expressed IgM^a, the IgH allotype associated with CH235 UCA V_H, and nearly 60% of B cells expressed IgM^b from the IgH^b allele (fig. S23C). Again, as with the DH270 KI model, the CH235 UCA V_H was presumably deleted through V_H replacement in IgM^{b+} B cells. With regard to light chain expression, the proportion of $Ig\lambda^+$ B cells (2%) was below normal. The profile suggests that the KI V_L inhibited the expression of endogenous mouse light chain and was not subject to active deletion by receptor editing. Together, these data suggest that expression of CH235 UCA is subject to negative selection at the heavy chain level, likely at the B cell progenitor stage, where V_H replacement takes place. B cells that have passed through this initial hurdle progress normally into mature B cells in peripheral lymphoid tissues.

Heterozygous ($V_H^{+/-}$, $V_L^{+/-}$) CH235 UCA KI mice were immunized five times with M5. G458Y-stabilized SOSIP trimer glycosylated with heterogeneous glycans or enriched for $Man_5GlcNAc_2$ glycans (Fig. 5B). In this model, approximately 10% of follicular B cells bind the CH505 M5.G458Y envelope, making it a relatively stringent model for testing engagement of bnAb precursors (fig. S24). $Man_5GlcNAc_2$-enriched and heterogeneously glycosylated M5.G458Y Env trimer elicited comparable serum neutralizing antibody titers (Fig. 5C). Serum neutralization could be knocked out by a N280D mutation indicative of a CD4bs epitope–targeting response (Fig. 5C). Serum antibodies from either group of trimer-immunized mice were capable of neutralizing the tier 2 autologous CH505 M5 virus (fig. S25B), which is resistant to the CH235 UCA, suggesting maturation of the CH235 antibody response (Fig. 5A).

By next-generation sequencing of the CH235 V_H region, we determined that M5 G458Y trimer-immunized mice generated CH235 sequences with the K19T improbable mutation (Fig. 5D). To determine the effects of immunogen affinity on K19T selection, we immunized CH235 UCA mice with low-affinity CH505 M5 gp120 that bound to the UCA with

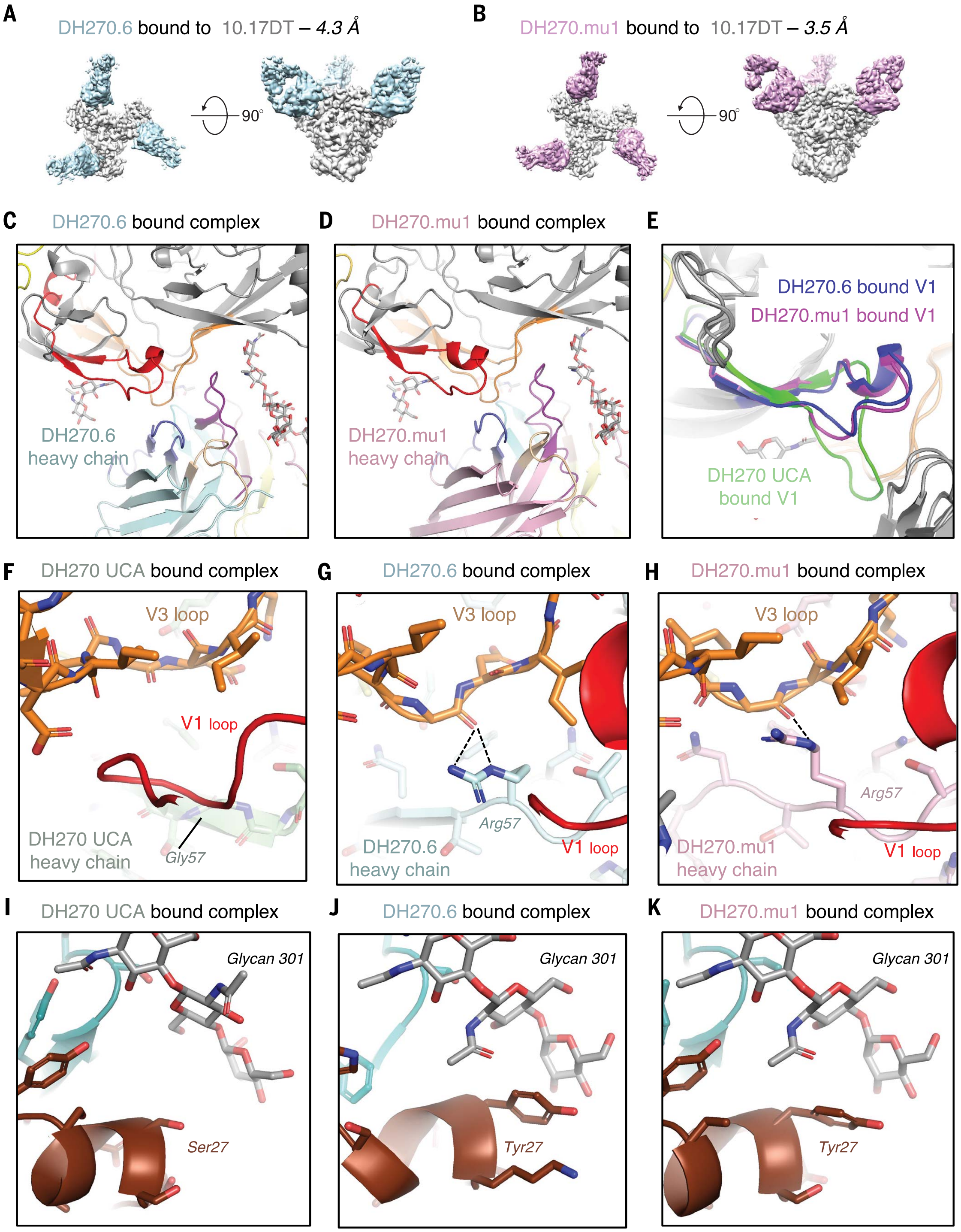


Fig. 4. Vaccine-induced antibody DH270.mu1 with improbable mutations binds to the HIV-1 Env SOSIP trimer with the same binding mode as DH270 bnAbs. (**A**) Cryo-EM reconstruction (4.3 Å resolution) of DH270.6 (cyan) bound to CH848 10.17DT–stabilized Env trimers (gray) segmented by components. (**B**) Same as (A) but for the DH270.mu1 complex (3.5 Å resolution) with DH270.mu1 shown in pink. (**C**) Zoomed-in view of the interactive region between DH270.6 and CH848 10.17DT. HIV-1 Env gp120 is shown in gray with the V3 loop in orange and the V1 loop in red. Glycans are shown in stick representation and colored by elements, with oxygen atoms in red and nitrogen atoms in blue. The heavy chain of DH270.6 is shown in cyan and light chain in yellow. Color code for CDRs: wheat, CDR H1; blue, CDR H2; purple, CDR H3; brown, CDR L1; pink, CDR L2; teal, CDR L3. (**D**) Same as (C) but for the DH270.mu1 complex, with DH270.mu1 heavy chain in pink. (**E**) Overlay of the DH270-bound V1 (green), DH270.6-bound V1 (blue), and DH270.mu1-bound V1 (magenta). (**F** to **H**) Zoomed-in views showing details of the contacts made by heavy chain residue 57. Black dashed lines indicate hydrogen bonds. (**I** to **K**) Zoomed-in views showing details of the contacts made by light chain residue 27.

4 μM affinity (fig. S25A). The highest-affinity immunogen elicited the highest frequency of K19T-encoding antibodies, and the lowest-affinity immunogen elicited the lowest frequency of K19T-encoding antibodies (Fig. 5, D and E). To determine the function of K19T-encoding antibodies, we isolated 57 CH235 monoclonal antibodies by sorting antigen-specific B cells from M5.G458Y trimer-immunized mice, 86% of which were somatically mutated (Fig. 5F). Three antibodies (termed CH235.mu1, 2, and 3) possessed the K19T mutation (Fig. 5G and figs. S25 and S26). In addition to the K19T mutation, CH235.mu1 acquired the improbable somatic mutation D99N, which is also

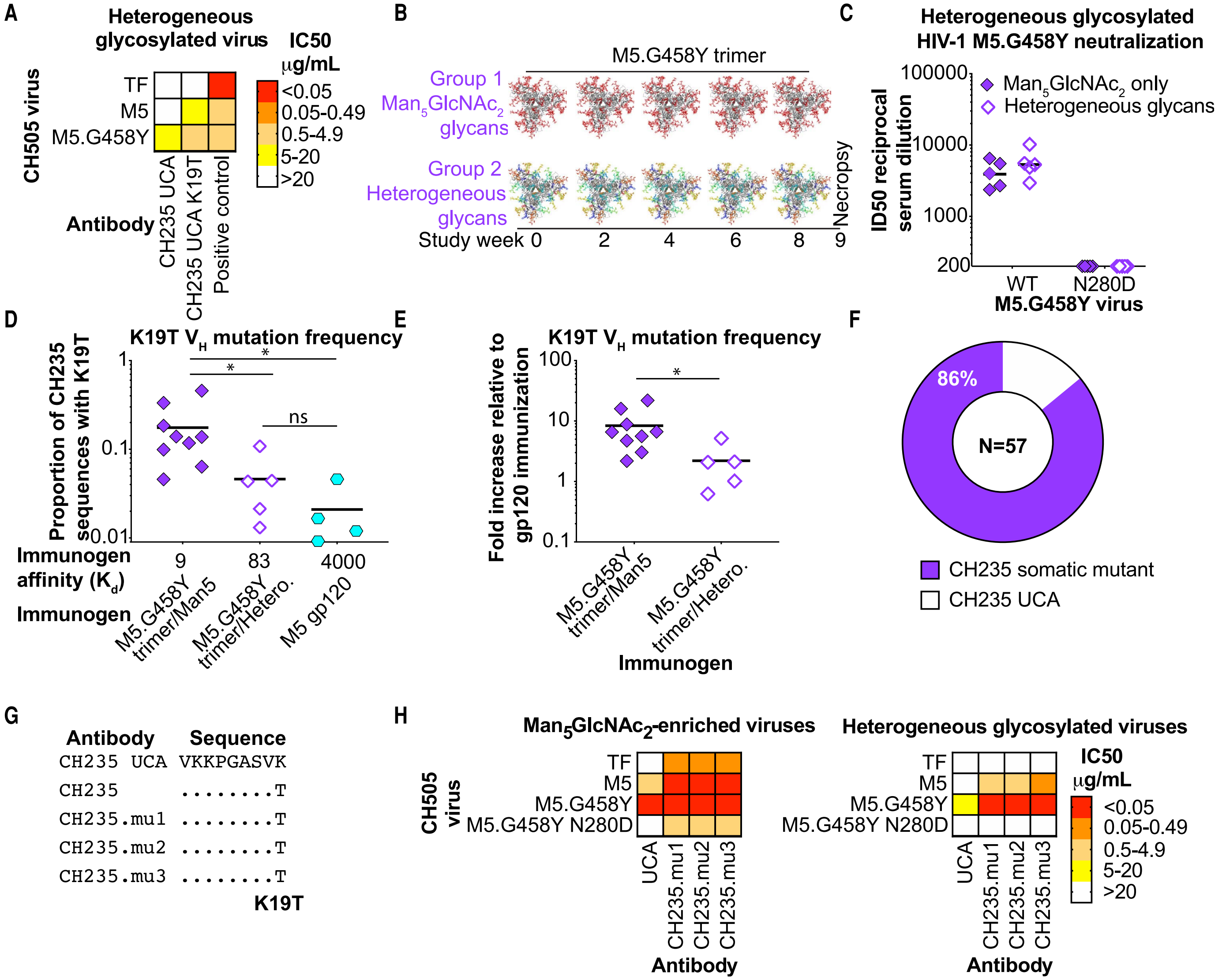


Fig. 5. Env trimer immunization elicits neutralizing antibodies with the critical, improbable K19T amino acid change in the CD4 binding site bnAb lineage CH235. (**A**) CH235 UCA neutralization of autologous CH505 virus infection of TZM-bl cells is enhanced by the acquisition of the K19T somatic mutation in the V_H. Neutralization titers are shown as IC_{50} values. A mixture of CH01 and CH31 was used as a positive control. (**B**) CH235 UCA (het/het) KI mouse (*n* = 5) immunization comparing $Man_5GlcNAc_2$-enriched versus heterogeneously glycosylated CH505 M5.G458Y SOSIP gp140 trimer. (**C**) Comparable titers of serum antibody neutralization of CH505 M5.G458Y virus infection of TZM-bl cells were elicited by $Man_5GlcNAc_2$-enriched versus heterogeneously glycosylated CH505 M5.G458Y SOSIP gp140 trimer. Neutralization activity was sensitive to a N280D amino acid change in the CD4 binding site. Neutralization titers are shown as ID_{50} with each symbol representing one mouse serum sample collected 1 week after the fourth immunization. Group geometric means are shown by horizontal bars. Murine leukemia virus was negative at each time point. (**D**) Enumeration by V_H next-generation sequencing of the frequency of unique CH235 sequences encoding the K19T amino acid change in mice immunized with M5 gp120 or $Man_5GlcNAc_2$-enriched M5.G458Y trimer. Frequencies of K19T were determined in total splenocytes 1 week after the final immunization. Each mouse is shown by an individual symbol; horizontal bars indicate the group mean. **P* < 0.05 (Wilcoxon exact test); ns, not significant. (**E**) Fold increase in K19T frequency in heterozygous CH235 UCA KI mice immunized with $Man_5GlcNAc_2$-enriched M5.G458Y trimer and M5 gp120. Symbols and bars are the same as in (D). **P* < 0.05 (Wilcoxon exact test). (**F**) M5.G458Y gp120-reactive single splenic B cells from a $Man_5GlcNAc_2$-enriched CH505 M5.G458Y Env trimer–immunized mouse were sorted by fluorescence-activated cell sorting (FACS) 1 week after the final immunization. All of the 57 recovered antibody sequences originated from the CH235 KI variable regions. The pie chart shows the percentage of CH235-expressing B cells that have acquired one or more somatic mutations. (**G**) Amino acid alignment of V_H sequences from vaccine-induced and infection-induced CH235 antibodies shows the occurrence of the K19T mutation. (**H**) Vaccine-induced CH235 antibodies encoding K19T neutralize autologous CH505 virus infection of TZM-bl cells more potently than does the CH235 UCA. Viruses were grown under conditions that result in $Man_5GlcNAc_2$ enrichment (left) or heterogeneous glycosylation (right). The heat maps depict IC_{50} neutralization titers for each individual virus.

present in CH235 (figs. S25 and S26). These three vaccine-induced antibodies more potently neutralized $Man_5GlcNAc_2$-enriched or heterogeneously glycosylated M5.G458Y virus than did the UCA (Fig. 5H and fig. S25D). These antibodies were specific for the CD4bs, as judged by their sensitivity to a N280D Env mutation (Fig. 5H). Vaccine-induced antibodies also neutralized heterogeneously glycosylated CH505 M5 and $Man_5GlcNAc_2$-enriched CH505 TF, which were both resistant to CH235 UCA neutralization (Fig. 5H, right). Therefore, vaccination stimulated these somatically mutated antibodies to progress beyond the neutralization phenotype of the CH235 UCA. Thus, we have demonstrated that the design of UCA-targeting Envs that bind with high affinity to B cells bearing key improbable mutations can overcome rare mutations as a maturational bottleneck in both V3-glycan and CD4bs bnAb B cell lineages.

Induction of potent neutralizing CD4bs antibodies in macaques

We immunized rhesus macaques to determine whether $Man_5GlcNAc_2$-enriched M5.G458Y SOSIP gp140 could elicit CH235-like CD4 binding site antibodies in outbred primates with wild-type antibody repertoires. Given previous successes of Env nanoparticle immunogens in macaques (*15*, *39*) and in the DH270 UCA KI mouse studies performed here, we generated a nanoparticle displaying the $Man_5GlcNAc_2$-enriched M5.G458Y SOSIP gp140 (Fig. 6A) using the two-step conjugation method described above (fig. S11). Macaques were immunized three times with M5.G458Y SOSIP gp140 nanoparticles formulated in the adjuvant 3M-052 in stable emulsion (Fig. 6B). Plasma IgG binding antibodies to the $Man_5GlcNAc_2$-enriched M5.G458Y SOSIP gp140 were detectable after a single immunization and continued to increase with the two subsequent boosts (Fig. 6C). Serum autologous tier 2 neutralizing antibodies against $Man_5GlcNAc_2$-enriched M5.G458Y virus were detected after the second immunization and were boosted by the third immunization (Fig. 6D). In previous studies, the elicitation of autologous tier 2 neutralizing antibodies in all vaccinated animals has been difficult to achieve, requiring osmotic pumps and fractionated immunizations over short time intervals (*53*, *54*). However, $Man_5GlcNAc_2$-enriched M5.G458Y nanoparticles successfully elicited potent neutralizing antibodies in all four macaques with two monthly immunizations (Fig. 6, D to H). Neutralizing antibodies were not restricted to $Man_5GlcNAc_2$ enrichment, as the heterogeneously glycosylated M5.G458Y virus produced in 293T cells was potently neutralized by all four animals also (Fig. 6E).

Next, we examined whether the neutralizing antibodies in the macaque serum had the same neutralization signature as CH235 antibodies early in bnAb development. The CH235 UCA plus the K19T mutation potently neutralized M5.G458Y, weakly neutralized M5, and showed undetectable neutralization toward CH505 TF virus (Fig. 5A). The macaque serum showed a similar neutralization signature, as the CH235 UCA plus the K19T mutation in that the serum neutralized $Man_5GlcNAc_2$-enriched M5.G458Y potently, M5 moderately, and CH505 TF weakly (Fig. 6, F and G). To map the neutralizing antibodies to the same CD4bs epitope as CH235, we examined neutralization of M5.G458Y virus with a knockout mutation in the CD4bs at position 280 (N280D), a knockout mutation characteristic of CH235-class bnAb precursors (*52*). After two and three immunizations, all four macaques generated N280-dependent neutralizing antibodies (Fig. 6, F and G). As a group, the macaques showed a factor of 34 decrease (range: >9 to 114) in serum neutralization potency against $Man_5GlcNAc_2$-enriched M5.G458Y when the CH235 epitope was eliminated (Fig. 6, H and I). Thus, the CD4bs immunogen that is capable of selecting improbable mutations in the CH235 lineage elicited CD4bs serum neutralizing antibodies in primates with neutralization signatures similar to that of the CH235 bnAb.

Concluding remarks

Improbable somatic mutations in HIV-1 bnAb variable regions are often required for their neutralization activity and represent obstacles to the induction of bnAbs by vaccination (*10*). We have shown that immunogens can be designed to select for specific antibody nucleotide mutations in the setting of vaccination. Our immunogen design approach was to design an immunogen that could bind to the initial precursor of the bnAb lineage with sufficient affinity to activate it, and could bind with higher affinity to the affinity maturation intermediates with improbable mutations. Our results show that the immune system, when provided with an affinity gradient between two clonally related antibodies, can be manipulated to generate HIV-1 bnAb improbable mutations. In addition to the CH235 and DH270 bnAb lineages, we envision that this vaccine design strategy will be applicable to other HIV-1 bnAb lineages.

We succeeded in eliciting CD4bs neutralizing antibodies in primates with a neutralization phenotype similar to CH235 intermediate antibodies. The next step for vaccine design will be to guide these antibodies to develop neutralization breadth. Unlike bnAbs such as PGT121 or VRC01, the development of the two bnAb lineages investigated here is not limited by rare insertion or deletion events. Instead, these antibodies encode improbable nucleotide mutations needed for neutralization activity (*18*, *22*, *23*). Thus, we hypothesize that the elicitation of DH270 V3-glycan or CH235 CD4bs types of bnAbs will require a mutation-guided vaccine design approach that specifically selects for improbable mutations (*10*). Among the 26 bnAbs for which we analyzed the number of improbable mutations, CH235.12 had the most V_H improbable mutations (*10*). Although the CH505 M5 G458Y Env trimer selected for K19T at a sufficiently high frequency for isolation of monoclonal antibodies encoding this mutation, the functionally important W47L improbable mutation (*10*) was not elicited with the current immunogen. The absence of the W47L mutation in the 57 monoclonal antibodies demonstrated that immunogen engagement of the UCA is a required step to initiate affinity maturation of the antibody lineage, but engagement alone is not sufficient to select for all improbable mutations. Not all amino acids encoded by improbable nucleotide mutations are required for neutralization breadth. Thus, defining the amino acids required for neutralization breadth comparable to CH235.12 or DH270.6 would provide a simplified maturation pathway for vaccine-induced antibody responses. This functionally required subset of mutations should be inducible with a vaccine, given our observation that the 134 vaccine-induced single DH270 antibodies encoded 16 of 19 V_H mutations present in the DH270.1 mature bnAb—including the four improbable mutations of Arg^{19}, Arg^{57}, Thr^{97}, and Thr^{98} (Fig. 3G and figs. S18 and S20). All 16 mutations did not occur on the same antibody; rather, they were scattered among the 134 antibodies (figs. S16, S18, and S19). Without proper antigen selection, it will take multiple decades of vaccination to elicit a DH270-like bnAb. This timeline could be accelerated by the design of sequential immunogens that select for the required combination of functional antibody somatic mutations (*13*, *14*, *55*, *56*).

It is important to consider the status of glycans on 293 cell line–produced Envs that we have used as immunogens. In Go *et al.*, multiple stabilized Envs forms were produced in 293, CHO, or CD4 T cell lines and their site-specific glycans were profiled (*57*, *58*). Remarkable similarity was found among all SOSIP trimes regardless of the cell line used for Env expression (*57*, *58*). Comparison of virion Env and recombinant SOSIP trimer glycans has also demonstrated glycosylation similarities (*59*). Thus, both 293 cell–produced and CHO-produced stabilized trimers have glycosylation patterns of predominant high-mannose glycans similar to those found on native virion Env. Evaluation of the CH848 10.17DT immunogen used in this study was consistent with these published glycosylation patterns.

Our study emphasizes the difficult nature of induction of broadly neutralizing antibodies, in that the design of sequential immunogens

must be precise, and the simultaneous use of sequential immunogens for multiple bnAb types also runs the risk of leading some bnAb B cell lineages off track. Thus, multiple rounds of iterative immunogen design will be needed to develop complex sequential immunogens that can induce a polyclonal bnAb response.

Finally, this study demonstrates proof-of-concept for targeted selection of improbable mutations to improve antibody affinity maturation. Although HIV-1 was the focus of this study, this strategy of selection of specific antibody nucleotides by immunogen design can be applied to B cell lineages targeting other pathogens where guided affinity

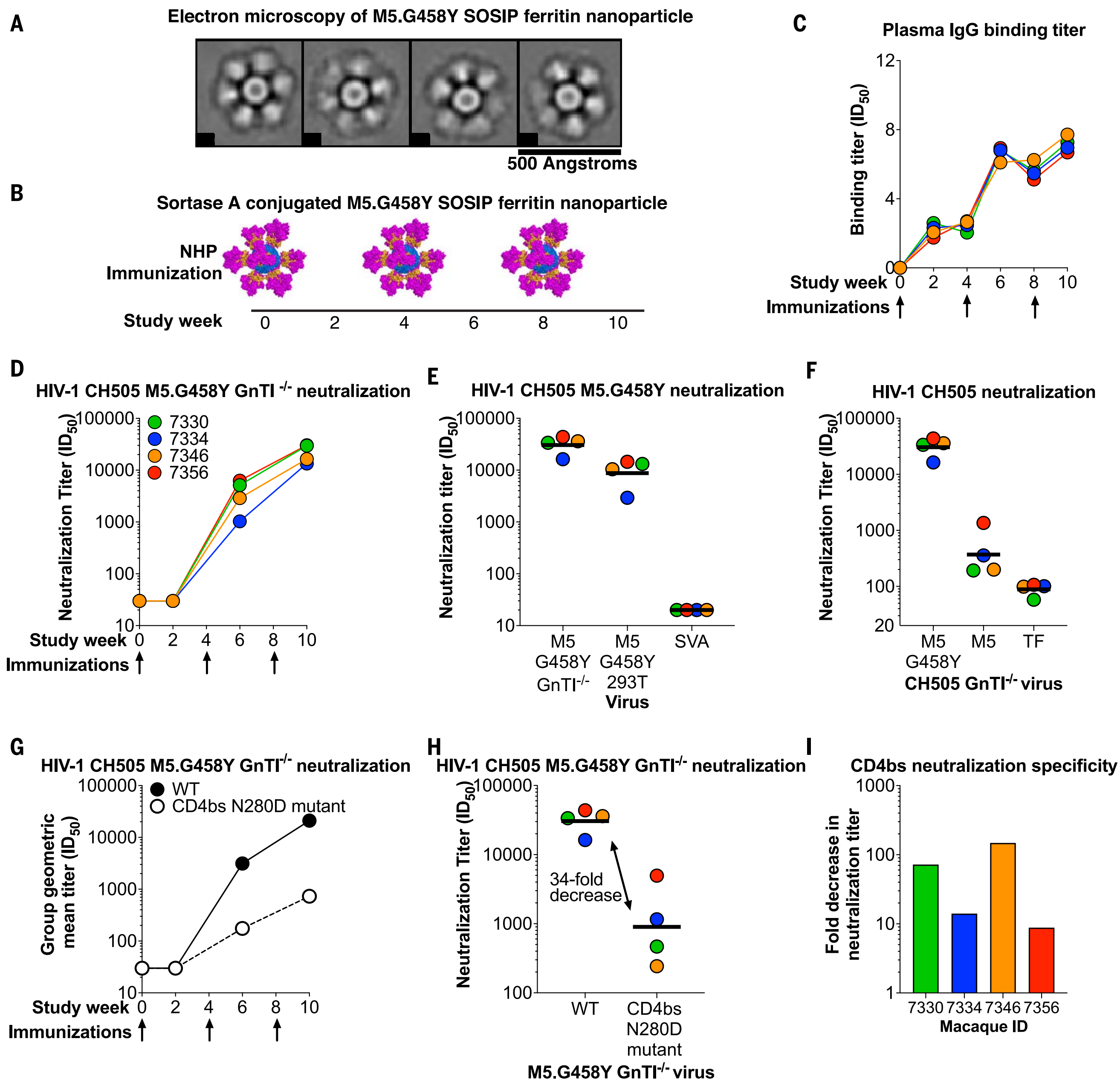


Fig. 6. CH505 M5.G458Y SOSIP nanoparticle immunization of rhesus macaques elicits serum autologous neutralizing antibodies against the CD4 binding site. (**A**) Two-dimensional class averages of negative-stain electron microscopy images of GnTI$^{-/-}$ cell–expressed M5.G458Y-stabilized SOSIP gp140 ferritin nanoparticles. (**B**) Rhesus macaque immunization with M5.G458Y-stabilized SOSIP gp140 ferritin nanoparticles enriched for $Man_5GlcNAc_2$ glycans. (**C**) Plasma IgG binding over time to GnTI$^{-/-}$ cell–expressed M5.G458Y-stabilized SOSIP gp140. (**D**) Serum neutralization kinetics against GnTI$^{-/-}$ cell–produced HIV-1 CH505 M5.G458Y virus. Each curve represents an individual macaque. (**E**) Macaque serum after three immunizations neutralizes $Man_5GlcNAc_2$-enriched and heterogeneously glycosylated CH505 M5.G458Y virus. Horizontal bars are the geometric mean for the group (n = 4). SVA is a negative control murine leukemia virus. (**F**) Macaque serum after three immunizations neutralizes $Man_5GlcNAc_2$-enriched CH505 M5.G458Y, M5, and TF viruses. Horizontal bars are the geometric mean for the group (n = 4). (**G**) Geometric mean of the neutralization titers of all four animals against CH505 M5.G458Y wild-type and CD4bs CH235 bnAb knockout (N280D) viruses over time. (**H**) Individual macaque neutralization titers after three immunizations against M5.G458Y wild-type and CD4bs CH235 bnAb knockout (N280D) viruses. Horizontal bars are the geometric mean for the group (n = 4). (**I**) Fold decrease in week 10 neutralization ID_{50} titers shown in (H) for each macaque upon mutation of the CD4 binding site with a N280D mutation.

maturation is needed for a protective antibody response.

Methods

Animals and immunizations

DH270 UCA V_H and V_L KI mice were generated by introducing rearranged $V_H(D)J_H$ and $V_\lambda J_\lambda$ exons into mouse J_H and J_κ loci of an F_1 (129/Sv × C57BL/6) ES cell line, respectively. The IgH alleles from the two mouse strains are distinguishable by allotypic difference. The DH270 UCA $V_H(D)J_H$ exon was integrated via homologous recombination into the J_H region of the IgH allele (IgH^a) from 129/Sv mouse strain (fig. S7A). The IgH^b allele from C57BL/6 strain was not modified. The integration replaces a 2255-bp region containing mouse J_H1-4 (starting from 877 bp upstream of J_H1 to 57 bp downstream of J_H4) with an expression cassette consisting of a promoter upstream of mouse V_H81X segment and rearranged $V_H(D)J_H$ exon encoding the V_H of DH270 UCA. Similarly, the DH270 UCA $V_\lambda J_\lambda$ exon was introduced into the Jκ region of one Igκ allele. The other Igκ allele was not altered and the two Igκ alleles are not distinguishable by allotypic difference. The V_L KI replaces a 1762-bp region containing mouse Jκ1-5 (starting from 114 bp upstream of Jκ1 to 286 bp downstream of Jκ5) with an expression cassette consisting of a promoter upstream of mouse Vκ3-1 segment. Correct V_H and V_L KI have been verified with Southern blotting. The ES clones were injected into Rag2 deficient blastocysts to generate chimeric mice, which were subsequently bred with 129/Sv mice for germline transmission. Germline mice were used for the immunization experiments. The CH235 UCA V_H and V_L KI mice were generated in the same way.

DH270 UCA heterozygous heavy chain variable and light chain variable region double KI mice ($V_HDJ_H{}^{+/-}$, $V_\lambda J_\lambda{}^{+/-}$ KI strains) were immunized six times 2 weeks apart with 25 μg of protein immunogens with 5 μg of the TLR4 agonist-based, IDRI proprietary adjuvant system GLA-SE. Immunizations were performed via intramuscular injection (200 μl). Heterozygous CH235 UCA ($V_HDJ_H{}^{+/-}$, $V_\kappa J_\kappa{}^{+/-}$) KI mice were immunized similarly. Control groups were administered in parallel with 5 μg of GLA-SE (adjuvant only). Blood samples were collected 10 days after each immunization for immune profiling.

Indian-origin rhesus macaques were housed and treated in AAALAC-accredited institutions. The study protocol and all veterinarian procedures were approved by the Duke University IACUC and were performed based on standard operating procedures developed by Bioqual (Rockville, MD). Macaques were immunized intramuscularly in the quadriceps every 4 weeks. Each macaque was administered 100 μg of CH505 M5.G458Y SOSIP gp140 ferritin nanoparticles adjuvanted with 30 μg of 3M-052 stable emulsion. 3M-052 stable emulsion was formulated as described (60). Whole blood and serum were drawn on the day of vaccination and 2 weeks after each immunization to examine immune responses.

Recombinant antibody and Fab production

Antibody and Fab was produced as described (61, 62). Briefly, recombinant proteins were expressed in Expi293 cells (Life Technologies) by transient transfection with Expifectamine (Invitrogen). Five days after transfection, cell culture media was cleared of cells and protein A (ThermoFisher) or KappaSelect (GE Healthcare) affinity chromatography was used to purify IgG or Fab, respectively. Purified protein was buffer exchanged into PBS with successive rounds of centrifugation, filtered, and stored at −80°C.

Recombinant SOSIP envelope production

HIV-1 Env SOSIP gp140 trimers were designed as chimeric SOSIP trimers (62). CH848 SOSIP gp140 trimers were stabilized with a 201C-433C disulfide bond (24). CH505 SOSIPs were stabilized with E64K and A316W mutations (25). Freestyle 293 (Life Technologies) cells were cultured in Freestyle 293 media below 3 × 10^6 cells ml^{-1}. On the day of transfection, cells were diluted to 1.25 × 10^6 cells ml^{-1} with fresh media and 1 liter of cells was transfected with 293Fectin (Life Technologies) complexed with 650 μg of envelope-expressing DNA and 150 μg of furin-expressing plasmid DNA. Cells were cultured for 6 days in shaker flasks. Cell culture supernatant was cleared of cells by centrifugation for 30 min at 3500 rpm and subsequently 0.8-μm filtered. The cell-free supernatant was concentrated to less than 100 ml with a single-use tangential flow filtration cassette and 0.8-μm filtered again. Trimeric Env protein was purified with PGT145 affinity chromatography. PGT145 IgG1 antibody (100 mg) was conjugated to 10 ml of CnBr-activated sepharose FastFlow resin (GE Healthcare). Coupled resin was packed into a Tricorn column (GE Healthcare) and stored in PBS supplemented with 0.05% sodium azide. Cell-free supernatant was applied to the column at 2 ml/min in PBS supplemented with 0.05% sodium azide using an AKTA Pure (GE Healthcare). The column was washed, and protein was eluted off of the column with 3 M $MgCl_2$. The eluate was immediately diluted in 10 mM Tris pH 8, 0.2-μm filtered, and concentrated down to 2 ml for size-exclusion chromatography. To produce biotinylated CH0848 10.17DT SOSIP gp140s, the envelope sequence was expressed with a C-terminal avidin tag (AviTag: GLNDIFEAQKIEWHE). After antibody affinity chromatography and eluate concentration, the envelope was dialyzed for 1 hour in 10 mM Tris pH 8. Envelope was biotinylated with the BirA biotin-protein ligase standard reaction kit (Avidity). The ligation reaction was done by agitating 25 μM of SOSIP trimer at 900 rpm at 30°C for 5 hours. The biotinylated protein was then concentrated to 2 ml for size-exclusion chromatography. Size-exclusion chromatography was performed with a Superose6 16/600 column (GE Healthcare) in 10 mM Tris pH 8, 500 mM NaCl. Fractions containing trimeric HIV-1 Env protein were pooled together, sterile-filtered, snap-frozen, and stored at −80°C.

Deglycosylation of CH848 10.17DT SOSIP Envs

Env samples were either partially or fully deglycosylated depending on the type of analysis. For disulfide analysis, samples containing 10 μg of the CH848 10.17DT SOSIP Envs were alkylated with a 10-fold molar excess of 4-vinylpyridine in the dark for 1 hour at room temperature (RT) to cap any potential free cysteine residues. Alkylated Env samples were subsequently deglycosylated with 500 U of PNGase F in 100 μl of 50 mM ammonium citrate buffer (pH 6.5) for 1 week at 37°C. The fully deglycosylated and alkylated samples were digested overnight with trypsin (protein/enzyme ratio of 30:1) at 37°C. For glycosylation analysis, samples containing 25 μg of the CH848 10.17DT SOSIP Envs were incubated with Endo H (2.5 μl, ≥5 units ml^{-1}) for 48 hours at 37°C in 50 mM ammonium acetate buffer, pH 5.5. Following Endo-H treatment, Env samples were digested as described below.

Proteolytic digestion of CH848 10.17DT SOSIP Envs for glycosylation analysis

CH848 10.17DT SOSIP Envs (25 μg) were denatured with 7 M urea in 100 mM Tris buffer (pH 8.0), reduced at RT for 1 hour with TCEP (5 mM), and alkylated with 20 mM IAM at RT for another hour in the dark. The reduced and alkylated samples were buffer-exchanged using a 50-kDa MWCO filter (Millipore) prior to protease digestion. Digestion was performed using trypsin alone and a combination of trypsin and chymotrypsin at a 30:1 protein/enzyme ratio. Samples were incubated overnight at 37°C. The resulting Env digest was either directly analyzed or stored at −20°C until further analysis. To ensure reproducibility of the method, digestion was performed at least three times on different days with samples obtained from the same batch and analyzed with the same experimental procedure.

Chromatography and mass spectrometry

High-resolution LC/MS experiments were performed using an LTQ-Orbitrap Velos Pro (Thermo Scientific) mass spectrometer equipped with ETD coupled to an Acquity UPLC M-Class system (Waters). Mobile phases consisted of solvent A (99.9% deionized H_2O + 0.1% formic acid) and solvent B (99.9% CH_3CN + 0.1%

formic acid). A 5-μl aliquot of the sample (~1.5 μM) was injected onto C18 PepMap 300 column (300 μm i.d. × 15 cm, 300 Å, Thermo Fisher Scientific) at a flow rate of 4 μl/min. The following CH_3CN/H_2O multistep gradient was used: 3% B for 5 min, followed a linear increase to 40% B in 50 min, then a linear increase to 90% B in 15 min. The column was held at 97% B for 10 min before reequilibration. All mass spectrometric analysis was performed in the positive ion mode using data-dependent acquisition mode: The five most intense ions in the survey scan in the mass range 400 to 2000 m/z were selected for alternating CID and ETD in the LTQ linear ion trap using a normalized collision energy of 30% for CID and an ion-ion reaction time of 100 to 150 ms for ETD. Full MS scans were measured at a resolution (R) of 30,000 at m/z 400. Under these conditions, the measured R (FWHM) in the orbitrap mass analyzer is 20,000 at m/z 1000 and 17,000 at m/z 1500.

Glycopeptide identification and disulfide bond analysis

Details of the glycopeptide compositional analysis have been described (63–65). Briefly, compositional analysis of glycopeptides was carried out by first identifying the peptide portion from tandem MS data. Once the peptide portion was determined, plausible glycopeptide compositions were obtained using the high-resolution MS data and GlycoPep DB. The putative glycopeptide composition was confirmed manually from CID and ETD data.

Disulfide bond patterns of CH848 10.17DT SOSIP Envs were determined by mapping the disulfide-linked peptides. Data analysis was performed using Mascot (v 2.5.1) search engine for peptides containing free cysteine residues and disulfide bond patterns were analyzed manually as described (66, 67).

In vitro HIV-1 neutralization

Antibody-mediated HIV-1 neutralization was measured using Tat-regulated luciferase (Luc) reporter gene expression to quantify reductions in virus replication in TZM-bl cells as described (68). TZM-bl cells were obtained from the NIH AIDS Research and Reference Reagent Program, as contributed by J. Kappes and X. Wu. The monoclonal antibody or serum was pre-incubated with virus (~150,000 relative light unit equivalents) for 1 hour at 37°C, and TZM-bl cells were subsequently added. After 48 hours, cells were lysed and Luc activity determined using a microtiter plate luminometer and BriteLite Plus Reagent (Perkin Elmer). Neutralization titers are the antibody concentration in μg/ml or serum reciprocal dilution at which relative luminescence units (RLU) were reduced by 50% compared to RLU in virus control wells after subtraction of background RLU in cell control wells (IC_{50} and ID_{50}, respectively).

Surface plasmon resonance (SPR)

SPR experiments were performed on a BIACore T200. To determine apparent affinities, approximately 300 RU (range 310 to 321 RU) of each antibody was captured on an anti-human IgFc immobilized Series S CM5 sensor chip (GE Healthcare). Serial dilutions of SOSIP Env was flowed over immobilized antibody in HEPES-buffered saline. To determine binding affinity, biotinylated SOSIP gp140 was immobilized on streptavidin-coated sensor chips. Serial dilutions of antibody Fab were flowed over the Env. Each concentration of Env was flowed over each immobilized antibody for 120 s and dissociation was measured for 600 s. In between injections of each Env concentration, the surface was regenerated by injecting glycine pH 2 for 30 s. Binding rate constants (k_a, k_d) were measured after global curve fitting to a Langmuir model. Curve fitting analysis was performed with BiaEvaluation software (GE Healthcare) using a 1:1 Langmuir model, or a heterogeneous binding model when appropriate to derive rate (k_a, k_d) and apparent or true equilibrium dissociation constants (K_d).

Serum IgG ELISA binding titers

Avi-tag SOSIP capture or PGT151 SOSIP capture ELISAs were performed as described (69). Protein (2 μg ml^{-1}) in sodium bicarbonate buffer was incubated in sealed Nunc-absorp (ThermoFisher) plates overnight at 4°C. Unbound protein was washed away and the plates were blocked with SuperBlock for 1 hour. Serial dilution of serum or monoclonal antibodies were added to the plate for 90 min. Binding antibodies were detected with HRP-labeled anti-IgG Fc antibody. HRP was detected with 3,3′,5,5′-tetramethylbenzidine. Binding titers were analyzed as area-under-curve of the log-transformed concentrations.

Serum IgG and monoclonal antibody competition ELISAs

Serum IgG or monoclonal antibody competition assays were performed as described (61, 70). In brief, Nuncsorp plates were coated with HIV-1 envelope, washed and blocked as stated above for direct ELISAs. After blocking was complete, mouse serum was diluted in SuperBlock at a 1:50 dilution and incubated in triplicate wells for 90 min. For monoclonal antibody competitors, a dilution series starting at 100 μg ml^{-1} was incubated in triplicate wells for 90 min. Non-biotinylated monoclonal antibodies were incubated with the Env in triplicate as positive controls for blocking. To determine relative binding, no plasma or no antibody was added to a group of wells scattered throughout the plate. After 90 min, the non-biotinylated antibody or plasma was washed away, and biotinylated monoclonal antibodies was incubated in the wells for 1 hour at sub-saturating concentrations. Each well was washed, and binding of biotinylated monoclonal antibodies was determined with a 1:8000 dilution of horseradish peroxidase (HRP)–conjugated streptavidin. HRP was detected with tetramethylbenzidine and stopped with 1% HCl. The absorbance at 450 nm of each well was read with a Spectramax plate reader (Molecular Devices). Binding of the biotinylated monoclonal antibody to HIV-1 envelope in the absence of plasma was compared to in the presence of plasma to calculate percent inhibition of binding. Based on historical negative controls, assays were considered valid if the positive control antibodies blocked greater than 20% of the biotinylated antibody binding.

Sortase A conjugation of HIV-1 envelope trimers to ferritin nanoparticles

CH848 10.17DT SOSIP gp140 was expressed with amino acids LPSTGG encoded at its C terminus. The CH848 10.17DT SOSIP was expressed in Freestyle293 cells and purified by affinity chromatography with broadly neutralizing trimer-specific antibody PGT151 as stated above. Trimeric gp140 was isolated by size-exclusion chromatography using a Superose6 16_60 column. Ferritin particles were expressed with a (GGGGG)3 repeat sequence encoded at the N terminus of each subunit. For affinity purification of ferritin particles, 6×His tags were appended to the C terminus. CH848 SOSIP with a C-terminal sortase A tag and ferritin particles with a sortase A N-terminal tag were buffer-exchanged into 50 mM Tris, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.5. 120 μM SOSIP gp140 was mixed with 120 μM ferritin subunits and incubated with 100 μM sortase A overnight at room temperature. After incubation, conjugated particles were isolated from free ferritin or free SOSIP gp140 by size-exclusion chromatography using a Superose6 16_60 column.

Antibody binding to HIV-1 envelope by biolayer interferometry (BLI)

Biolayer interferometry was performed as described (71). Antibodies were immobilized to anti-Fc sensor tips and incubated with 50 μg ml^{-1} of Env SOSIP gp140-nanoparticle or SOSIP trimer solutions for 400 s. To compare binding profiles between Env SOSIP gp140 and Env SOSIP gp140 nanoparticles, BLI binding values were normalized to PGT151 binding response by dividing each antibody binding response value by the PGT151 binding response value.

Antigen-specific fluorescence-activated single B cell sorting

To sort Env-specific memory B cells, splenocytes were incubated with the following

fluorochrome-antibody conjugates (all from BD Biosciences): FITC anti-mouse IgG1 (A85-1), FITC anti-mouse IgG2a/2b (R2-40), FITC anti-mouse IgG3 (R40-82), PE anti-mouse GL7 (GL7), PE-Cy7 anti-mouse IgM (R6-60.2), AlexaFluor700 anti-mouse CD19 (1D3), BV510 anti-mouse IgD (11-26C.2a), and BV650 anti-mouse B220 (RA3-6B2). Cells were also labeled with BV421- and AF647-conjugated SOSIP trimers (CH848 10.17) to identify Env-specific memory B cells. Env-specific memory B cells were identified as viable $B220^{+}CD19^{+}IgM^{-}IgD^{-}GL7^{-}IgG1/2/3^{+}$ cells that bound both BV421- and AF647-conjugated SOSIP trimers. In some cases, the BV421-SOSIP contained the N332A mutation; in these samples Env-specific memory B cells were identified as those binding the AF647-CH848 10.17 wild-type SOSIP but not the BV421-N332A SOSIP mutant. Single cells were sorted on a BD FACS Aria II into wells of 96-well PCR plates containing lysis buffer. Plates were immediately frozen and stored at –80°C.

Antibody cloning by PCR

Ig genes were amplified as described, with some modifications (*72*, *73*). Ig genes from a single B cell were reverse-transcribed with Superscript III (ThermoFisher) using random hexamer oligonucleotides as primers. The complementary DNA was used to perform nested PCR for heavy and light chain genes using AmpliTaq gold (ThermoFisher). In parallel, PCR reactions were done with mouse Ig-specific primers and DH270 variable region-specific primers. In one reaction, mouse variable region and mouse constant region primers were used. In a separate reaction, human variable region and mouse constant region primers were used. Positive PCR amplification of Ig genes was identified by gel electrophoresis. Positive PCR reactions were purified using the PCR clean-up kit (Qiagen). The purified PCR amplicon was sequenced with 4 µM of forward and reverse primers. Contigs of the PCR amplicon sequence were made, and genes were inferred with the human library and the mouse library in Cloanalyst (*74*). Antibody genes were categorized as human or mouse according to which species had the closest sequence identity. A second aliquot of the purified PCR amplicon was used for overlapping PCR to generate a linear expression cassette. The expression cassette was transfected with Effectene (Qiagen) into 293T cells. The supernatant containing recombinant antibodies was cleared of cells by centrifugation and used for binding assays. The genes of selected heavy chains were synthesized as IgG1 (GenScript). Kappa and lambda chains were synthesized similarly. Plasmids were prepared for transient transfection using the Megaprep plasmid plus kit (Qiagen).

B cell receptor–dependent calcium flux in Ramos B cells

Ramos B cell calcium flux was measured as described (*75*). Protein tetramers were formed at a 4:1 molar ratio of protein to streptavidin (Invitrogen). Ramos cell lines stably expressing DH270 UCA, DH270.1, or CH65 IgM (*76*) were passaged (1:10) 4 days before calcium flux experiments. On the day of the experiment, cells with >95% viability were resuspended at 10^6 cells ml^{-1} in 2:1 ratio of RPMI media (GIBCO) + FLIPR Calcium 6 dye (Molecular Devices). Cells were plated in a U-bottom 96-well tissue culture plate (Costar) and incubated at 37°C 5% CO_2 for 2 hours. In a black clear-bottom 96-well plate (Costar) containing 50 µl of RPMI media (GIBCO) + FLIPR Calcium 6 dye (Molecular Devices) (2:1 ratio) either 0.1 nmol of proteins or 50 µg ml^{-1} of anti-human IgM F(ab′)2 (Jackson Immuno) were added (based on a 100-µl volume). Using a FlexStation 3 multimode microplate reader (Molecular Devices), 50 µl of supernatant containing cells were transferred into the 50 µl of media containing protein or anti-human IgM F(ab′)2 (Jackson Immuno) and continuously read for 5 min. Relative fluorescent value units were background-subtracted and the data expressed as percentage of the IgM maximum signal (% IgM max).

B cell receptor–dependent calcium flux in mouse B cells

Calcium flux in murine B cells was evaluated as described (*56*). For experiments evaluating total splenic B cells, WT BL/6 and DH270 UCA double V_HDJ_H homozygous and V_LJ_L homozygous KI splenocytes were collected, and total B cells were enriched using a mouse Pan-B cell isolation kit (Stemcell) according to manufacturer's instructions. Enriched Pan-B cells were stained by LIVE/DEAD Fixable Yellow Dead Cell Stain Kit (ThermoFisher Scientific) for 30 min. For experiments evaluating splenic B cell subsets, single-cell suspensions were directly stained with cell surface marker combinations for spleen subfractionation into transitional-B and mature-B subsets including 0.5 µg ml^{-1} of anti-B220-BV786 (catalog #563894), anti-CD19 APCR700 (catalog #565473), and anti-CD93 BV650 (catalog #563807). For both sets of experiments, prestained B cells were loaded with Fluo-4 via thorough washes in HBSS, followed by mixing with equal volumes of 2× Fluo-4 Direct calcium reagent loading solution (Fluo-4 Direct Calcium Assay Kits, ThermoFisher Scientific). After sequential 30-min incubations at 37°C and room temperature, cells were washed and incubated with LIVE/DEAD staining buffer for 30 min and resuspended in calcium-containing HBSS and incubated at room temperature for 5 min, before being activated by 25 µg ml^{-1} anti-IgM $F(ab')_2$ (Southern Biotech). Fluo-4 data for B cells were acquired on a BD LSR II flow cytometer and analyzed by FlowJo software.

T cell and B cell phenotyping of murine splenocytes

To analyze B cell phenotypes in the DH270 UCA and CH235 UCA $V_H + V_L$ KI mice, spleens were dissected from the mice and dissociated by grinding into cell suspension by mechanical disruption with a syringe plunger on a 70-µm cell strainer. Single-cell suspensions of mouse spleens were prepared then treated with ammonium chloride lysis solution or with red blood cell lysing buffer (Sigma R7757) to lyse red blood cells. Cells were counted on a Countess (ThermoFisher). For enumeration of germinal center B cells, cells were incubated with optimal concentrations of the following fluorochrome-antibody conjugates (all from BD Biosciences): PE anti-mouse GL7 (GL7), PE-CF594 anti-mouse CD93 (AA4.1), AlexaFluor700 anti-mouse CD19 (1D3), BV605 anti-mouse CD95 (Jo2), BV650 anti-mouse B220 (RA3-6B2), and BV711 anti-mouse CD138 (281-2). Cells were subsequently labeled with Live/Dead Fixable Near-IR Dead Cell Stain (ThermoFisher) to allow exclusion of dead cells from analysis. Germinal center B cells were identified as viable $CD138^{-}B220^{+}CD19^{+}CD93^{-}GL7^{+}CD95^{+}$ cells. To enumerate Tfh cells, splenocytes were stained with the following antibody conjugates (all from BD Biosciences or Biolegend): FITC anti-mouse CD4 (RM4-5), PE anti-mouse CD25 (7D4), PE-CF594 anti-mouse PD1 (J43), PE-Cy5 anti-mouse TER119 (TER119), PE-Cy7 anti-mouse CD62L, Biotin anti-mouse CXCR5 (2G8), AlexaFluor700 anti-mouse CD8 (53-6.7), BV421 anti-mouse CD127 (SB/199), BV510 anti-mouse CD3 (145-2C11), BV570 anti-mouse CD11b (M1/70), BV650 anti-mouse NK1.1 (PK136), BV711 anti-mouse CD44 (IM7), and BV786 anti-mouse B220 (RA3-6B2). Tfh cells were identified as viable $TER119^{-}B220^{-}NK1.1^{-}CD3^{+}CD4^{+}CD8^{-}CD62L^{-}CD44^{+}CD25^{-}PD1^{+}CXCR5^{+}$ cells. After cell labeling, cells were fixed in 2% formaldehyde. Cells were acquired on a BD LSRII cytometer and analyzed with FlowJo version 10 software. To generate the data shown in figs. S7 and S23, splenocytes were stained with the following antibodies: APC anti- B220 (eBioscience 17-0452-83), PE anti-Thy1.2 (PharMingen 553006), PE anti-IgM (eBioscience 12-5790-83), PE anti-IgM^a (PharMingen 553517), PE anti-Igλ (Biolegend 407308), FITC anti-IgD (BD PharMingen 553439), FITC anti-IgM^b (PharMingen 553520), FITC anti-Igκ (1050-02) and viability dye, Sytox blue (Life Technologies S34857). The stained cells were analyzed on Attune NxT flow cytometer from Invitrogen and the data were analyzed with FlowJo10 software. During the analysis, a lymphocyte gate was drawn on FSC and SSC plot. Within this population, Sytox blue–negative live cells were gated. Single cells were gated on the next FSC-H and FSC-A plot. The

staining pattern of these single cells are shown in figs. S7A and S23A. B220[+] B cells were further gated on single cells, and the staining patterns of B220[+] B cells are shown in figs. S7, B to D, and S23, B to D.

Cryo-EM data collection and processing

To prepare Env complexes, CH848 10.17DT. SOSIP trimer at a final concentration of 1 mg ml^{-1} was incubated with 4- to 6-fold molar excess of the DH270 Fab fragments for 30 to 60 min. To prevent aggregation during vitrification, the sample was incubated in 0.085 mM dodecyl-maltoside (DDM). The specimen was vitrified by applying 2.5 μl of sample to freshly plasma-cleaned Quantifoil R 1.2/1.3 300-mesh Cu holey carbon grids, allowing the sample to adsorb to the grid for 60 s, followed by blotting with filter paper and plunge-freezing into liquid ethane using the Leica EM GP cryo-plunger (Leica Microsystems) (20°C, >90% relative humidity). For the DH270 UCA and DH270.6 complexes, data were acquired using the EPU automated data-acquisition program. Images were collected on a Titan Krios (Thermo Fisher) operating at 300 keV equipped with a Falcon III direct electron detector operating in counting mode. For the DH270 UCA and DH270.6 complexes, 3383 and 2844 movies, respectively, were collected at a magnification of 75,000× with a physical pixel size of 1.08 Å per pixel using a nominal defocus range of –1.25 to –3 μm. Each movie (30 frames) was acquired using a dose rate of ~0.8 e$^-$/pixel/s and a total exposure of 42 e$^-$/Å^2.

For the DH270.mu1 complex, data were acquired using the Gatan Latitude data collection software installed on a Titan Krios electron microscope operating at 300 kV and fitted with a Gatan K3 direct detection device operating in counting mode. We collected 3009 movies at a nominal magnification of 22,500× with a physical pixel size of 1.07 Å per pixel using a nominal defocus range of –1.25 to –3 μm. Each movie (60 frames) was acquired using a dose rate of ~1.01 e$^-$/pixel/s and a total exposure of 60.6 e$^-$/Å^2.

For the DH270 UCA and DH270.6 complexes, motion correction and dose weighting were performed using MotionCor2 (*77*). For the DH270.mu1 complex, motion correction and dose weighting were performed using Unblur (*78*). CTF was estimated using CTFFIND4 (*79*). Particles were picked using the Laplacian-of-Gaussian function in RELION-3 (*80*). These particles were imported into cryoSparc v2 (*81*), 2D classification was performed, and selected 2D classes representing different views of the complex were used for template-based particle picking in cryoSparc. Following further 2D classifications to remove junk, ab initio reconstruction and classification was performed using C1 symmetry. A 3D class was identified with three antibody Fabs bound symmetrically to the HIV-1 Env trimer. This initial model was refined using C3 symmetry against the clean stack of particles. Overall map resolution was reported according to the FSC$_{0.143}$ gold-standard criterion (*82*).

Cryo-EM model fitting

Fits of HIV-1 trimer and Fab to the cryo-EM reconstructed maps were performed using Chimera (www.rbvi.ucsf.edu/chimera) (*83*). BG505 SOSIP trimer structure (PDB ID 5YFL) was used for the trimer fits and the coordinates of DH270 UCA3 (PDB ID 5U15) and DH270.6 (PDB ID 5TQA) were used for fitting the Fab in the DH270 UCA and DH270.6 complex structures, respectively. The sequences were replaced with those of the CH848 10.17DT SOSIP trimer using Coot (*84*). The coordinates were further fit to the electron density first using Rosetta (*85*), followed by an iterative process of manual fitting using Coot and real-space refinement within Phenix (*86*). Molprobity (*87*) and EMRinger (*88*) were used to check geometry and evaluate structures at each iteration step. Figures were generated in UCSF Chimera and PyMOL (PyMOL Molecular Graphics System, Version 2.0; Schrödinger LLC). Map-fitting cross correlations were calculated using Fit-in-Map feature in UCSF Chimera. Local resolution of cryo-EM maps was determined using RELION.

Negative-stain electron microscopy of HIV-1 envelope

Electron microscopy was performed as described (*62*).

Differential scanning calorimetry

Envelope thermal denaturation profiles were determined as described (*89*). Envelope profiles were generated in HEPES-buffered saline (HBS; 10 mM HEPES, 150 mM NaCl pH 7.4) at concentrations ranging from 0.2 to 0.4 mg ml^{-1} using the NanoDSC platform (TA instruments, New Castle, DE). The observed, irreversible denaturation profiles were buffer-subtracted, converted to molar heat capacity, baseline-corrected with a sixth-order polynomial, and fit with three Gaussian transition models using the NanoAnalyze software (TA Instruments). The primary transition temperature (T_m) is reported as the temperature at the maximum observed heat capacity.

High-throughput heavy chain variable region sequencing

RNA was extracted from total splenocytes post-sixth immunization using the Qiagen RNeasy Mini isolation kit (Qiagen) and used for reverse transcription with random hexamer primers for cDNA synthesis. After cDNA synthesis, the DH270 or CH235 UCA KI Ig genes were amplified by PCR using a forward primer that anneals in the IGHV1 leader sequence and two reverse primers that anneal in the mouse IgG1/2 and IgG3 constant region sequences to amplify all human encoding IGHV1 IgG sequences. All primers have leading sequences that match Illumina adapter sequences for Nextera amplification. A second PCR step was performed to add Nextera index sequencing adapters (Illumina) and libraries were purified and size-selected by AMPpure bead cleanup. Libraries were quantified by quantitative PCR using the KAPA SYBR FAST qPCR kit (KAPA Biosystems) and sequenced using the Illumina Miseq V2 2× 300-bp kit.

Antibody sequence analysis

NGS reads from immunized mice were assembled using FLASh (*90*), quality filtered using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and deduplicated and aligned to their respective UCA sequence using in-house–developed bioinformatics programs. NGS reads were immunogenetically annotated using Cloanalyst (*91*) and sequences that were deemed nonfunctional (out-of-frame, missing invariant residues or CDRs) were discarded from analysis. Probability of mutations was estimated using the ARMADiLLO (Antigen Receptor Mutation Analyzer for Detecting Low Likelihood Occurrences) program (*10*). Briefly, given a UCA sequence and the number of mutations observed in the antibody sequence of interest, ARMADiLLO simulates somatic hypermutation based on a model of AID targeting and base substitution (*92*) and uses these simulations to estimate the probability of an amino acid at a specific position.

Mathematical modeling of improbable mutation acquisition with vaccination

We model the acquisition of improbable mutations as a Poisson process with rate parameter, λ, in which the average number of improbable mutations is 1 in a time interval T of 12 weeks based on the number of improbable mutations observed for the mice immunized biweekly in this study. For simplicity, here we assume improbable mutations are acquired at a constant rate in the absence of targeted selection during vaccination, although evolutionary rates of B cell lineages are known to fluctuate during infection (*49*). The probability of k or fewer Poisson-distributed events occurring within a time interval is a function of the cumulative density of the Poisson distribution:

$$P(X \le k) = \exp(-\lambda)\sum_{i=0}^{k}\frac{\lambda^i}{i!}$$

The Poisson cumulative distribution function can be expressed using the regularized incomplete gamma function Q as

$$P(X \le k) = Q[(k-1),\lambda] = \frac{\Gamma[(k-1),\lambda]}{\Gamma(\lambda)}$$

The probability of k or more events occurring in the time interval is then

$$P(X \geq k) = 1 - Q(k, \lambda)$$

which can be solved for λ numerically using the inverse incomplete regularized gamma function,

$$\lambda = Q^{-1}[1 - P(X \geq k), k]$$

The total time to acquire k or more mutations is then λT.

The probability of acquiring at least $k = 4$ improbable mutations given a rate parameter of 1 improbable mutation every 12 weeks is plotted as a function of time in fig. S19C. For $P(X \geq 4) = 0.99$, the number of 12-week intervals is 10.05 to acquire at least four improbable mutations. Thus, mice would need to be immunized biweekly for ~120 weeks in order to achieve 99% probability of acquiring at least four improbable mutations.

Quantification and statistical analysis

The statistical analyses for this paper were performed in SAS 9.4 to calculate exact Wilcoxon tests for group comparisons. Due to the exploratory nature of this research and the small sample size, we are using an alpha level of 0.05 as a descriptive level for significance and have not made any adjustments to control for multiple testing. For group sizes less than 5, no paired-sample comparisons were performed due to the small sample size; only descriptive statistics are provided in these instances.

REFERENCES AND NOTES

1. B. F. Haynes, D. R. Burton, Developing an HIV vaccine. *Science* **355**, 1129–1130 (2017). doi: 10.1126/science.aan0662; pmid: 28302812
2. B. F. Haynes, J. R. Mascola, The quest for an antibody-based HIV vaccine. *Immunol. Rev.* **275**, 5–10 (2017). doi: 10.1111/imr.12517; pmid: 28133795
3. B. F. Haynes *et al*., Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308**, 1906–1908 (2005). doi: 10.1126/science.1111781; pmid: 15860590
4. B. F. Haynes *et al*., HIV-Host Interactions: Implications for Vaccine Design. *Cell Host Microbe* **19**, 292–303 (2016). doi: 10.1016/j.chom.2016.02.002; pmid: 26922989
5. F. Klein *et al*., Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153**, 126–138 (2013). doi: 10.1016/j.cell.2013.03.018; pmid: 23540694
6. H. X. Liao *et al*., Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013). doi: 10.1038/nature12053; pmid: 23552890
7. X. Wu *et al*., Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection. *Cell* **161**, 470–485 (2015). doi: 10.1016/j.cell.2015.03.004; pmid: 25865483
8. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007). doi: 10.1146/annurev.biochem.76.061705.090740; pmid: 17328676
9. A. G. Betz, C. Rada, R. Pannell, C. Milstein, M. S. Neuberger, Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: Clustering, polarity, and specific hot spots. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2385–2388 (1993). doi: 10.1073/pnas.90.6.2385; pmid: 8460148
10. K. Wiehe *et al*., Functional Relevance of Improbable Antibody Mutations for HIV Broadly Neutralizing Antibody Development. *Cell Host Microbe* **23**, 759–765.e6 (2018). doi: 10.1016/j.chom.2018.04.018; pmid: 29861171
11. C. A. Schramm, D. C. Douek, Beyond Hot Spots: Biases in Antibody Somatic Hypermutation and Implications for Vaccine Design. *Front. Immunol.* **9**, 1876 (2018). doi: 10.3389/fimmu.2018.01876; pmid: 30154794
12. J. K. Hwang *et al*., Sequence intrinsic somatic mutation mechanisms contribute to affinity maturation of VRC01-class HIV-1 broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8614–8619 (2017). doi: 10.1073/pnas.1709203114; pmid: 28747530
13. B. F. Haynes, G. Kelsoe, S. C. Harrison, T. B. Kepler, B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat. Biotechnol.* **30**, 423–433 (2012). doi: 10.1038/nbt.2197; pmid: 22565972
14. A. Escolano *et al*., Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. *Cell* **166**, 1445–1458.e12 (2016). doi: 10.1016/j.cell.2016.07.030; pmid: 27610569
15. A. Escolano *et al*., Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* **570**, 468–473 (2019). doi: 10.1038/s41586-019-1250-z; pmid: 31142836
16. R. N. Germain, The art of the probable: System control in the adaptive immune system. *Science* **293**, 240–245 (2001). doi: 10.1126/science.1062946; pmid: 11452112
17. A. Ribas, J. D. Wolchok, Cancer immunotherapy using checkpoint blockade. *Science* **359**, 1350–1355 (2018). doi: 10.1126/science.aar4060; pmid: 29567705
18. M. Bonsignori *et al*., Staged induction of HIV-1 glycan-dependent broadly neutralizing antibodies. *Sci. Transl. Med.* **9**, eaai7514 (2017). doi: 10.1126/scitranslmed.aai7514; pmid: 28298420
19. L. M. Walker *et al*., Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011). doi: 10.1038/nature10373; pmid: 21849977
20. C. N. Daniels, K. O. Saunders, Antibody responses to the HIV-1 envelope high mannose patch. *Adv. Immunol.* **143**, 11–73 (2019). doi: 10.1016/bs.ai.2019.08.002; pmid: 31607367
21. L. Kong *et al*., Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat. Struct. Mol. Biol.* **20**, 796–803 (2013). doi: 10.1038/nsmb.2594; pmid: 23708606
22. F. Gao *et al*., Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell* **158**, 481–491 (2014). doi: 10.1016/j.cell.2014.06.022; pmid: 25065977
23. M. Bonsignori *et al*., Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. *Cell* **165**, 449–463 (2016). doi: 10.1016/j.cell.2016.02.022; pmid: 26949186
24. Y. D. Kwon *et al*., Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol.* **22**, 522–531 (2015). doi: 10.1038/nsmb.3051; pmid: 26098315
25. S. W. de Taeye *et al*., Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* **163**, 1702–1715 (2015). doi: 10.1016/j.cell.2015.11.056; pmid: 26687358
26. A. Torrents de la Peña *et al*., Improving the Immunogenicity of Native-like HIV-1 Envelope Trimers by Hyperstabilization. *Cell Rep.* **20**, 1805–1817 (2017). doi: 10.1016/j.celrep.2017.07.077; pmid: 28834745
27. D. Jung, C. Giallourakis, R. Mostoslavsky, F. W. Alt, Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* **24**, 541–570 (2006). doi: 10.1146/annurev.immunol.23.021704.115830; pmid: 16551259
28. L. R. Covey, P. Ferrier, F. W. Alt, VH to VHDJH rearrangement is mediated by the internal VH heptamer. *Int. Immunol.* **2**, 579–583 (1990). doi: 10.1093/intimm/2.6.579; pmid: 2085492
29. R. Kleinfield *et al*., Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1+ B-cell lymphoma. *Nature* **322**, 843–846 (1986). doi: 10.1038/322843a0; pmid: 3092106
30. R. W. Kleinfield, M. G. Weigert, Analysis of VH gene replacement events in a B cell lymphoma. *J. Immunol.* **142**, 4475–4482 (1989). pmid: 2498430
31. C. Chen, Z. Nagy, E. L. Prak, M. Weigert, Immunoglobulin heavy chain gene replacement: A mechanism of receptor editing. *Immunity* **3**, 747–755 (1995). doi: 10.1016/1074-7613(95)90064-0; pmid: 8777720
32. S. L. Tiegs, D. M. Russell, D. Nemazee, Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* **177**, 1009–1020 (1993). doi: 10.1084/jem.177.4.1009; pmid: 8459201
33. D. Gay, T. Saunders, S. Camper, M. Weigert, Receptor editing: An approach by autoreactive B cells to escape tolerance. *J. Exp. Med.* **177**, 999–1008 (1993). doi: 10.1084/jem.177.4.999; pmid: 8459227
34. E. L. Prak, M. Weigert, Light chain replacement: A new model for antibody gene rearrangement. *J. Exp. Med.* **182**, 541–548 (1995). doi: 10.1084/jem.182.2.541; pmid: 7629511
35. G. Kelsoe, L. Verkoczy, B. F. Haynes, Immune System Regulation in the Induction of Broadly Neutralizing HIV-1 Antibodies. *Vaccines* **2**, 1–14 (2014). doi: 10.3390/vaccines2010001; pmid: 24932410
36. L. Verkoczy *et al*., Autoreactivity in an HIV-1 broadly reactive neutralizing antibody variable region heavy chain induces immunologic tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 181–186 (2010). doi: 10.1073/pnas.0912914107; pmid: 20018688
37. C. Doyle-Cooper *et al*., Immune tolerance negatively regulates B cells in knock-in mice expressing broadly neutralizing HIV antibody 4E10. *J. Immunol.* **191**, 3186–3191 (2013). doi: 10.4049/jimmunol.1301285; pmid: 23940276
38. Y. Chen *et al*., Common tolerance mechanisms, but distinct cross-reactivities associated with gp41 and lipids, limit production of HIV-1 broad neutralizing antibodies 2F5 and 4E10. *J. Immunol.* **191**, 1260–1275 (2013). doi: 10.4049/jimmunol.1300770; pmid: 23825311
39. P. Martinez-Murillo *et al*., Particulate Array of Well-Ordered HIV Clade C Env Trimers Elicits Neutralizing Antibodies that Display a Unique V2 Cap Approach. *Immunity* **46**, 804–817.e7 (2017). doi: 10.1016/j.immuni.2017.04.021; pmid: 28514687
40. C. D. Morris *et al*., Differential Antibody Responses to Conserved HIV-1 Neutralizing Epitopes in the Context of Multivalent Scaffolds and Native-Like gp140 Trimers. *mBio* **8**, e00036-17 (2017). doi: 10.1128/mBio.00036-17; pmid: 28246356
41. J. Ingale *et al*., High-Density Array of Well-Ordered HIV-1 Spikes on Synthetic Liposomal Nanoparticles Efficiently Activate B Cells. *Cell Rep.* **15**, 1986–1999 (2016). doi: 10.1016/j.celrep.2016.04.078; pmid: 27210756
42. L. He *et al*., Presenting native-like trimeric HIV-1 antigens with self-assembling nanoparticles. *Nat. Commun.* **7**, 12041 (2016). doi: 10.1038/ncomms12041; pmid: 27349934
43. M. Kanekiyo *et al*., Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102–106 (2013). doi: 10.1038/nature12202; pmid: 23698367
44. F. D. Batista, M. S. Neuberger, B cells extract and present immobilized antigen: Implications for affinity discrimination. *EMBO J.* **19**, 513–520 (2000). doi: 10.1093/emboj/19.4.513; pmid: 10675320
45. K. Sliepen *et al*., Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **12**, 82 (2015). doi: 10.1186/s12977-015-0210-4; pmid: 26410741
46. T. Tokatlian *et al*., Innate immune recognition of glycans targets HIV nanoparticle immunogens to germinal centers. *Science* **363**, 649–654 (2019). doi: 10.1126/science.aat9120; pmid: 30573546
47. D. Fera *et al*., HIV envelope V3 region mimic embodies key features of a broadly neutralizing antibody lineage epitope. *Nat. Commun.* **9**, 1111 (2018). doi: 10.1038/s41467-018-03565-6; pmid: 29549260
48. S. Dutta, P. Sengupta, Men and mice: Relating their ages. *Life Sci.* **152**, 244–248 (2016). doi: 10.1016/j.lfs.2015.10.025; pmid: 26596563
49. Z. Sheng *et al*., Effects of Darwinian Selection and Mutability on Rate of Broadly Neutralizing Antibody Evolution during HIV-1 Infection. *PLOS Comput. Biol.* **12**, e1004940 (2016). doi: 10.1371/journal.pcbi.1004940; pmid: 27191167
50. M. Tian *et al*., Induction of HIV Neutralizing Antibody Lineages in Mice with Diverse Precursor Repertoires. *Cell* **166**, 1471–1484.e18 (2016). doi: 10.1016/j.cell.2016.07.029; pmid: 27610571
51. R. K. Abbott *et al*., Precursor Frequency and Affinity Determine B Cell Competitive Fitness in Germinal Centers, Tested with Germline-Targeting HIV Vaccine Immunogens. *Immunity* **48**, 133–146.e6 (2018). doi: 10.1016/j.immuni.2017.11.023; pmid: 29287996
52. C. C. LaBranche *et al*., Neutralization-guided design of HIV-1 envelope trimers with high affinity for the unmutated common ancestor of CH235 lineage CD4bs broadly neutralizing antibodies. *PLOS Pathog.* **15**, e1008026 (2019). doi: 10.1371/journal.ppat.1008026; pmid: 31527908
53. K. M. Cirelli *et al*., Slow Delivery Immunization Enhances HIV Neutralizing Antibody and Germinal Center Responses via Modulation of Immunodominance. *Cell* **177**, 1153–1171.e28 (2019). doi: 10.1016/j.cell.2019.04.012; pmid: 31080066
54. M. Pauthner *et al*., Elicitation of Robust Tier 2 Neutralizing Antibody Responses in Nonhuman Primates by HIV Envelope

Trimer Immunization Using Optimized Approaches. *Immunity* **46**, 1073–1088.e6 (2017). doi: 10.1016/j.immuni.2017.05.007; pmid: 28636956
55. D. C. Malherbe *et al*., Sequential immunization with a subtype B HIV-1 envelope quasispecies partially mimics the in vivo development of neutralizing antibodies. *J. Virol.* **85**, 5262–5274 (2011). doi: 10.1128/JVI.02419-10; pmid: 21430056
56. W. B. Williams *et al*., Initiation of HIV neutralizing B cell lineages with sequential envelope immunizations. *Nat. Commun.* **8**, 1732 (2017). doi: 10.1038/s41467-017-01336-3; pmid: 29170366
57. E. P. Go *et al*., Comparative Analysis of the Glycosylation Profiles of Membrane-Anchored HIV-1 Envelope Glycoprotein Trimers and Soluble gp140. *J. Virol.* **89**, 8245–8257 (2015). doi: 10.1128/JVI.00628-15; pmid: 26018173
58. E. P. Go *et al*., Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers. *J. Virol.* **91**, e02428-16 (2017). doi: 10.1128/JVI.02428-16; pmid: 28202756
59. W. B. Struwe *et al*., Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. *Cell Rep.* **24**, 1958–1966.e5 (2018). doi: 10.1016/j.celrep.2018.07.080; pmid: 30134158
60. N. Van Hoeven *et al*., A Formulated TLR7/8 Agonist is a Flexible, Highly Potent and Effective Adjuvant for Pandemic Influenza Vaccines. *Sci. Rep.* **7**, 46426 (2017). doi: 10.1038/srep46426; pmid: 28429728
61. K. O. Saunders *et al*., Vaccine Elicitation of High Mannose-Dependent Neutralizing Antibodies against the V3-Glycan Broadly Neutralizing Epitope in Nonhuman Primates. *Cell Rep.* **18**, 2175–2188 (2017). doi: 10.1016/j.celrep.2017.02.003; pmid: 28249163
62. K. O. Saunders *et al*., Vaccine Induction of Heterologous Tier 2 HIV-1 Neutralizing Antibodies in Animal Models. *Cell Rep.* **21**, 3681–3690 (2017). doi: 10.1016/j.celrep.2017.12.028; pmid: 29281818
63. E. P. Go *et al*., Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J. Proteome Res.* **8**, 4231–4242 (2009). doi: 10.1021/pr9002728; pmid: 19610667
64. E. P. Go *et al*., Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome Res.* **7**, 1660–1674 (2008). doi: 10.1021/pr7006957; pmid: 18330979
65. J. Irungu *et al*., Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass Spectrom.* **19**, 1209–1220 (2008). doi: 10.1016/j.jasms.2008.05.010; pmid: 18565761
66. E. P. Go, D. Hua, H. Desaire, Glycosylation and disulfide bond analysis of transiently and stably expressed clade C HIV-1 gp140 trimers in 293T cells identifies disulfide heterogeneity present in both proteins and differences in O-linked glycosylation. *J. Proteome Res.* **13**, 4012–4027 (2014). doi: 10.1021/pr5003643; pmid: 25026075
67. E. P. Go, Y. Zhang, S. Menon, H. Desaire, Analysis of the disulfide bond arrangement of the HIV-1 envelope protein CON-S gp140 ΔCFI shows variability in the V1 and V2 regions. *J. Proteome Res.* **10**, 578–591 (2011). doi: 10.1021/pr100764a; pmid: 21114338
68. D. C. Montefiori, Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **485**, 395–405 (2009). doi: 10.1007/978-1-59745-170-3_26; pmid: 19020839
69. L. D. Williams *et al*., Potent and broad HIV-neutralizing antibodies in memory B cells and plasma. *Sci. Immunol.* **2**, eaal2200 (2017). doi: 10.1126/sciimmunol.aal2200; pmid: 28783671
70. S. L. Hulot *et al*., Comparison of Immunogenicity in Rhesus Macaques of Transmitted-Founder, HIV-1 Group M Consensus, and Trivalent Mosaic Envelope Vaccines Formulated as a DNA Prime, NYVAC, and Envelope Protein Boost. *J. Virol.* **89**, 6462–6480 (2015). doi: 10.1128/JVI.00383-15; pmid: 25855741
71. S. M. Alam *et al*., Mimicry of an HIV broadly neutralizing antibody epitope with a synthetic glycopeptide. *Sci. Transl. Med.* **9**, eaai7521 (2017). doi: 10.1126/scitranslmed.aai7521; pmid: 28298421
72. H. X. Liao *et al*., High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. *J. Virol. Methods* **158**, 171–179 (2009). doi: 10.1016/j.jviromet.2009.02.014; pmid: 19428587
73. R. Zhang *et al*., Initiation of immune tolerance-controlled HIV gp41 neutralizing B cell lineages. *Sci. Transl. Med.* **8**, 336ra62 (2016). doi: 10.1126/scitranslmed.aaf0618; pmid: 27122615
74. T. B. Kepler, Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000 Res.* **2**, 103 (2013). doi: 10.12688/f1000research.2-103.v1; pmid: 24555054
75. M. Bonsignori *et al*., Inference of the HIV-1 VRC01 Antibody Lineage Unmutated Common Ancestor Reveals Alternative Pathways to Overcome a Key Glycan Barrier. *Immunity* **49**, 1162–1174.e8 (2018). doi: 10.1016/j.immuni.2018.10.015; pmid: 30552024
76. G. C. Weaver *et al*., In vitro reconstitution of B cell receptor-antigen interactions to evaluate potential vaccine candidates. *Nat. Protoc.* **11**, 193–213 (2016). doi: 10.1038/nprot.2016.009; pmid: 26741406
77. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
78. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). doi: 10.7554/eLife.06980; pmid: 26023829
79. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015). doi: 10.1016/j.jsb.2015.08.008; pmid: 26278980
80. J. Zivanov *et al*., New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). doi: 10.7554/eLife.42166; pmid: 30412051
81. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
82. S. H. Scheres, S. Chen, Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012). doi: 10.1038/nmeth.2115; pmid: 22842542
83. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
84. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
85. F. DiMaio *et al*., Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015). doi: 10.1038/nmeth.3286; pmid: 25707030
86. P. D. Adams *et al*., PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002). doi: 10.1107/S0907444902016657; pmid: 12393927
87. V. B. Chen *et al*., MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010). doi: 10.1107/S0907444909042073; pmid: 20057044
88. B. A. Barad *et al*., EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015). doi: 10.1038/nmeth.3541; pmid: 26280328
89. R. Henderson *et al*., Selection of immunoglobulin elbow region mutations impacts interdomain conformational flexibility in HIV-1 broadly neutralizing antibodies. *Nat. Commun.* **10**, 654 (2019). doi: 10.1038/s41467-019-08415-7; pmid: 30737386
90. T. Magoč, S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011). doi: 10.1093/bioinformatics/btr507; pmid: 21903629
91. T. B. Kepler *et al*., Reconstructing a B-Cell Clonal Lineage. II. Mutation, Selection, and Affinity Maturation. *Front. Immunol.* **5**, 170 (2014). doi: 10.3389/fimmu.2014.00170; pmid: 24795717
92. G. Yaari *et al*., Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. *Front. Immunol.* **4**, 358 (2013). doi: 10.3389/fimmu.2013.00358; pmid: 24298272

ACKNOWLEDGMENTS

We thank C. Fox and S. Reed for formulation of 3M-052 adjuvant in stable emulsion; C. Bowman, G. Stephens, A. Newman, and C. Marsini for veterinary technical assistance; E. Carter, K. Anasti, M. Barr, C. Vivian, and A. Foulger for technical assistance with immunoassays; M. A. Moody, L. Armand, and D. Marshall for flow cytometric assistance; and G. Hernandez, K. Mansouri, A. Sanzone, E. Machiele, E. Lee, K. Tilahun, J. Smoot, P. Powers, and R. Reed for protein and DNA production assistance. Flow cytometry and FACS were performed in the Duke Human Vaccine Institute Flow Cytometry Shared Resource. Differential scanning calorimetry was performed by the Duke Human Vaccine Institute Biomolecular Interaction Analysis Facility. Cryo-EM data were collected at the Shared Materials Instrumentation Facility at Duke University as part of the Molecular Microscopy Consortium. Cryo-EM image quality was monitored on-the-fly during data collection using routines developed by A. Bartesaghi. **Funding:** This work was supported by NIAID extramural project grant R01-AI120801 (K.O.S.), NIH, NIAID, Division of AIDS UM1 grant AI100645 for the Center for HIV/AIDS Vaccine Immunology-Immunogen Discovery (CHAVI-ID; B.F.H.), NIH, NIAID, Division of AIDS UM1 grant AI144371 for the Consortium for HIV/AIDS Vaccine Development (CHAVD; B.F.H.), NIAID extramural project grant R01-AI125093 (H.D.), NIAID extramural project grant R01-AI087202 (L.K.V.), and funding from the Duke Translational Health Initiative (P.A.). This work was also supported by the US NIH Intramural Research Program, US National Institute of Environmental Health Sciences (ZIC ES103326; to M.J.B.), and the Howard Hughes Medical Institute (F.W.A.). The funders had no role in data collection and interpretation or the decision to submit the work for publication. **Author contributions:** Experimental design: K.O.S., K.W., P.A., T.B., S.M.A., F.W.A., M.T., B.F.H. Investigation and assays: A.E., M.T., P.A., B.W., T.B., H.C., X.L., C.J., J.Z., E.G., D.E., A.E., D.W.C., H.L.C., N.R.W., K.M., P.W., A.C.-W., R.S., L.S., D.C.M., M.J.B. Supervision: K.O.S., K.W., M.T., P.A., D.W.C., M.B., M.J.B., H.D., S.M.A., T.B., L.V., D.C.M., F.W.A., B.F.H., R.J.E., M.G.L., M.T., S.G.R. Data analysis: K.O.S., K.W., M.T., W.R., R.J.E., P.A., T.B., S.M.A., R.H., A.L.H., D.E., M.B., E.G., H.D., L.V., D.W.C., D.C.M., F.W.A., B.F.H. Writing: B.F.H., K.O.S., K.W., M.T., and P.A. with editing from all other co-authors. **Competing interests:** K.O.S. and B.F.H. are inventors on International Patent Application PCT/US2018/020788 submitted by Duke University, that covers the composition and use of CH848 HIV-1 envelopes for induction of HIV-1 antibodies. D.C.M., C.L., K.O.S., K.W., and B.F.H. are inventors on International Patent Application PCT/US2018/03477submitted by Duke University, that covers the composition and use of CH505 HIV-1 envelopes for induction of HIV-1 antibodies. **Data and materials availability:** Antibody sequences have been deposited to GenBank under accession numbers MN643173 through MN643554. The cryo-EM maps and refined coordinates were deposited in the EMDB and RCSB PDB databases, respectively, under the following accession numbers: DH270 UCA (EMD-20817 and PDB ID 6UM5), DH270.6 (EMD-20818 and PDB ID 6UM6), and DH270.mu1(EMD-20819 and PDB ID 6UM7). The ARMADiLLO program is available for download at http://sites.duke.edu/ARMADiLLO. All flow cytometry data are available upon request. All other data are in the main and supplementary figures and text. The DH270UCA V_H/V_L KI mice and CH235UCA V_H/V_L KI mice are available from F.W.A.'s laboratory under a standard material transfer agreement with Boston Children's Hospital. 3M-052 stable emulsion adjuvant is available from M. Tomai and S. Reed under a material transfer agreement with 3M Company (St. Paul, MN) and Infectious Disease Research Institute (Seattle, WA), respectively.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaay7199/suppl/DC1
Figs. S1 to S26
Table S1
Reference (*93*)

View/request a protocol for this paper from *Bio-protocol*.

12 July 2019; accepted 5 November 2019
10.1126/science.aay7199

RESEARCH ARTICLE SUMMARY

HIV VACCINES

A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses

Jon M. Steichen*, Ying-Cing Lin*, Colin Havenar-Daughton*, Simone Pecetta*, Gabriel Ozorowski*, Jordan R. Willis*, Laura Toy, Devin Sok, Alessia Liguori, Sven Kratochvil, Jonathan L. Torres, Oleksandr Kalyuzhniy, Eleonora Melzi, Daniel W. Kulp, Sebastian Raemisch, Xiaozhen Hu, Steffen M. Bernard, Erik Georgeson, Nicole Phelps, Yumiko Adachi, Michael Kubitz, Elise Landais, Jeffrey Umotoy, Amanda Robinson, Bryan Briney, Ian A. Wilson, Dennis R. Burton, Andrew B. Ward, Shane Crotty†, Facundo D. Batista†, William R. Schief†

INTRODUCTION: HIV newly infects 1.8 million people each year, making development of an HIV vaccine a global health priority. Nearly all licensed vaccines protect by inducing antibodies, but highly variable pathogens such as HIV and influenza virus have eluded traditional vaccine strategies. The discoveries of broadly neutralizing antibodies (bnAbs) that bind to conserved epitopes on the surface proteins of these viruses have inspired vaccine design strategies to induce bnAbs. Antibodies are produced by B cells, and highly effective antibodies like bnAbs acquire affinity-enhancing mutations when a bnAb-precursor B cell mutates and matures from the original naïve B cell (or germline) state. Among several new vaccine strategies, germline-targeting vaccine design aims to induce bnAbs by first stimulating bnAb-precursor B cells and then shepherding B cell affinity maturation with a series of rationally designed boosting immunogens. A key rationale for this strategy is that germline-reverted forms of bnAbs—precursors with all recognizable amino acid mutations reverted to germline—typically have no detectable affinity for HIV envelope (Env). Thus, for a vaccine to initiate bnAb induction, a germline-targeting priming immunogen with appreciable affinity for bnAb precursors must be engineered.

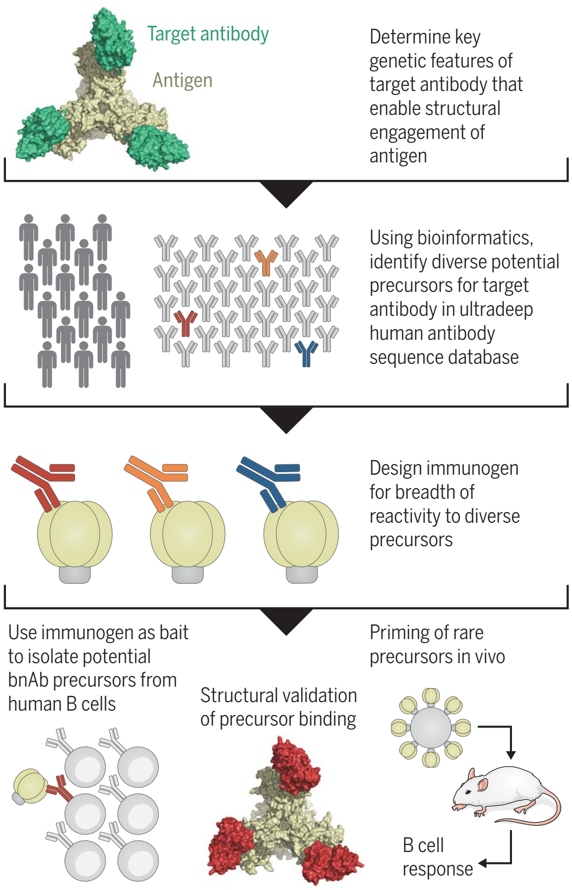
RATIONALE: Most HIV bnAbs (and most antibodies to any pathogen) bind to their target by using their heavy chain complementarity-determining region 3 (HCDR3) as a major binding determinant. Hence, an optimal HIV vaccine that induces multiple bnAbs, and a general solution to germline-targeting vaccine design that could be applied broadly to other pathogens, will need to work with HCDR3-dependent antibodies. However, the need to design germline-targeting immunogens to initiate HCDR3-dependent bnAb responses faces major technical challenges. Although each B cell expresses a single unique antibody, different B cells produce diverse antibodies encoded by different combinations of antibody genes, with the greatest antibody genetic diversity encoded in the HCDR3 portion of the molecule. The exceptional diversity in the human B cell repertoire makes any single HCDR3 sequence an impractical vaccine target. Rather, a pool of precursors sharing a set of bnAb-associated genetic features must be identified and targeted. Thus, owing to the enormous diversity of human antibodies, a germline-targeting immunogen should have affinity for diverse bnAb precursors in order to succeed in diverse vaccine recipients.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax4380

RESULTS: Herein we report a solution to the above challenges. Using the strongly HCDR3-dependent bnAb BG18 that binds a conserved site on HIV Env as a high-value target and a proof of principle, we demonstrate a method to identify pools of potential bnAb precursors in an ultradeep human antibody sequence database, guided by key genetic features that enable bnAb structural recognition of the antigen. We then use a representative set of those potential bnAb precursors as design targets to guide our engineering of HIV Env immunogens that bind to diverse potential bnAb precursors. Lastly, we provide critical preclinical validation of immunogen design by assessing these immunogens for (i) their ability to select rare potential bnAb-precursor naïve B cells from the blood of healthy human donors, (ii) their modes of binding to bnAb precursors, and (iii) their capacity to prime rare bnAb-precursor B cells with physiologically relevant affinities in a mouse model.

CONCLUSION: Overall, we demonstrate a new approach to defining diverse precursors for a target antibody and designing vaccine immunogens that take advantage of that information. The approach lays out a generalizable pathway for the development and preclinical validation of germline-targeting immunogens to stimulate precursors for HCDR3-dependent antibodies. ■



General strategy for germline-targeting vaccine design. Four key steps are defined here for design and validation of germline-targeting immunogens: structural and genetic understanding of target antibody-antigen interaction, identification of diverse antibody precursors, design of an immunogen to bind diverse precursors, and preclinical immunogen validation by isolating human B cell binders, assessing structural interaction with precursors, and stimulating responses in transgenic mice.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: schief@scripps.edu (W.R.S.); fbatista1@mgh.harvard.edu (F.D.B.); shane@lji.org (S.C.)
Cite this article as J. M. Steichen *et al.*, *Science* **366**, eaax4380 (2019). DOI: 10.1126/science.aax4380

RESEARCH ARTICLE

HIV VACCINES

A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses

Jon M. Steichen[1,2,3]*, Ying-Cing Lin[4]*, Colin Havenar-Daughton[3,5]*, Simone Pecetta[4]*, Gabriel Ozorowski[2,3,6]*, Jordan R. Willis[1,2,3]*, Laura Toy[3,5], Devin Sok[1,2,3], Alessia Liguori[1,2,3], Sven Kratochvil[4], Jonathan L. Torres[2,3,6], Oleksandr Kalyuzhniy[1,2,3], Eleonora Melzi[4], Daniel W. Kulp[1,2,3,7], Sebastian Raemisch[1,2,3], Xiaozhen Hu[1,2,3], Steffen M. Bernard[2,3,6], Erik Georgeson[1,2,3], Nicole Phelps[1,2,3], Yumiko Adachi[1,2,3], Michael Kubitz[1,2,3], Elise Landais[1,2,3], Jeffrey Umotoy[1,2,3], Amanda Robinson[1,2,3], Bryan Briney[1,2,3,8], Ian A. Wilson[2,3,6,9], Dennis R. Burton[1,2,3], Andrew B. Ward[2,3,6], Shane Crotty[3,5,10]†, Facundo D. Batista[4,11]†, William R. Schief[1,2,3,4]†

Vaccine induction of broadly neutralizing antibodies (bnAbs) to HIV remains a major challenge. Germline-targeting immunogens hold promise for initiating the induction of certain bnAb classes; yet for most bnAbs, a strong dependence on antibody heavy chain complementarity-determining region 3 (HCDR3) is a major barrier. Exploiting ultradeep human antibody sequencing data, we identified a diverse set of potential antibody precursors for a bnAb with dominant HCDR3 contacts. We then developed HIV envelope trimer–based immunogens that primed responses from rare bnAb-precursor B cells in a mouse model and bound a range of potential bnAb-precursor human naïve B cells in ex vivo screens. Our repertoire-guided germline-targeting approach provides a framework for priming the induction of many HIV bnAbs and could be applied to most HCDR3-dominant antibodies from other pathogens.

HIV infects 1.8 million new people each year, making development of an HIV vaccine a global health priority (*1*). Nearly all licensed vaccines protect by inducing antibodies, but highly antigenically variable pathogens such as HIV and influenza virus have eluded traditional vaccine strategies (*2*, *3*). The discoveries of broadly neutralizing antibodies (bnAbs) that bind to relatively conserved epitopes on viral surface proteins have inspired new vaccine design strategies (*4*, *5*).

Antibodies are produced by B cells and acquire affinity-enhancing mutations when the B cell mutates and matures from the original naïve (or germline) state. Germline-targeting HIV vaccine design aims to induce bnAbs by first priming bnAb-precursor B cells and then shepherding B cell affinity maturation with a series of rationally designed boosting immunogens. A key rationale for this strategy is that germline-reverted forms of bnAbs—precursors with all recognizable amino acid mutations reverted to germline—typically have no detectable affinity for HIV envelope (Env) proteins. Thus, for a vaccine to initiate bnAb induction, a germline-targeting priming immunogen with appreciable affinity for bnAb precursors must be engineered. Most HIV bnAbs (and most antibodies to any pathogen) bind to their target by using their heavy chain complementarity-determining region 3 (HCDR3) as a major binding determinant. Hence, an optimal HIV vaccine that induces multiple bnAbs to different HIV Env sites, and a general solution to germline-targeting vaccine design that could be applied broadly to other pathogens, will need to work with HCDR3-dependent antibodies. Many advances have been made in developing germline-targeting immunogens to prime precursors for one particular class of bnAbs (i.e., VRC01-class bnAbs) (*6*–*15*), and at least one such immunogen has entered human clinical testing (*16*). However, VRC01-class bnAbs represent a specialized case in which non-HCDR3 features are the main determinants of antibody specificity and affinity (*6*–*15*).

The need to design germline-targeting immunogens to initiate HCDR3-dependent bnAb responses brings new challenges. Although each B cell expresses a single unique antibody, different B cells produce diverse antibodies encoded by different combinations of antibody genes, with additional variation at junctions between genes, and the greatest antibody diversity is encoded in the HCDR3 portion of the molecule. The exceptional diversity in the human B cell repertoire makes any single bnAb-precursor HCDR3 sequence an impractical vaccine target. Rather, a pool of precursors sharing a set of bnAb-associated genetic features must be identified and targeted. Thus, owing to the antibody diversity in humans, a germline-targeting immunogen should have affinity for diverse bnAb precursors in order to succeed in diverse vaccine recipients.

Strategy for immunogen design and testing

We report a potential solution to the above challenges. We selected the bnAb BG18 (*17*, *18*) as a test case for a high-value vaccine design target, because BG18 is the most potent bnAb directed to the Asn[332] (N332) supersite, one of the major bnAb sites on HIV Env, and BG18 lacks insertions or deletions (indels) and therefore may be easier to induce than other bnAbs that require indels (see the supplementary materials) (*19*). Using the strongly HCDR3-dependent bnAb BG18 (*17*, *18*), we demonstrate a method to identify pools of bnAb potential precursors and use them as design targets to engineer HIV Env trimer immunogens that bind diverse bnAb potential precursors. We then provide preclinical validation by assessing these immunogens for (i) their ability to select rare bnAb potential precursor naïve B cells from the blood of HIV-seronegative human donors, (ii) their modes of binding to bnAb precursors, and (iii) their capacity to prime rare bnAb naïve precursors with human physiological affinities in a mouse model (fig. S1).

Precursor frequency analysis

Crystal structures of BG18 bound to HIV Env trimers indicated a BG18 binding mode in which the HCDR3 engages the conserved Gly-Asp-Ile-Arg (GDIR) motif at the base of the V3 loop like the bnAb PGT121 while the HCDR1 contacts the relatively conserved N332 glycan and the light chain (LC) straddles the V1 loop of gp120, unlike PGT121 (*18*). This binding mode was corroborated by (i) structural modeling (fig. S2, A to D); (ii) a 4.4-Å resolution cryo–electron microscopy (cryo-EM) structure of BG18 bound to an HIV Env trimer (Fig. 1A, fig. S3, and table S2); (iii) mutagenesis studies (fig. S2, E to F); and (iv) structural model-guided design of a minimally mutated BG18 bnAb (minBG18) that retained ~67% of the neutralization breadth of BG18 with only 11% amino acid mutations in the variable (V) gene regions of immunoglobulin heavy and light chains (V_H and V_L) compared with ~30% for BG18 (fig. S4). The successful design of minBG18 provided an additional rationale for BG18

[1]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA. [3]Consortium for HIV/AIDS Vaccine Development, The Scripps Research Institute, La Jolla, CA 92037, USA. [4]The Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02139, USA. [5]Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA 92037, USA. [6]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [7]Vaccine and Immune Therapy Center, The Wistar Institute, Philadelphia, PA 19104, USA. [8]Center for Viral Systems Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [9]Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [10]Division of Infectious Diseases, Department of Medicine, University of California, San Diego, La Jolla, CA 92037, USA. [11]Department of Immunology, Harvard Medical School, Boston, MA 02115, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: schief@scripps.edu (W.R.S.); fbatista1@mgh.harvard.edu (F.D.B.); shane@lji.org (S.C.)

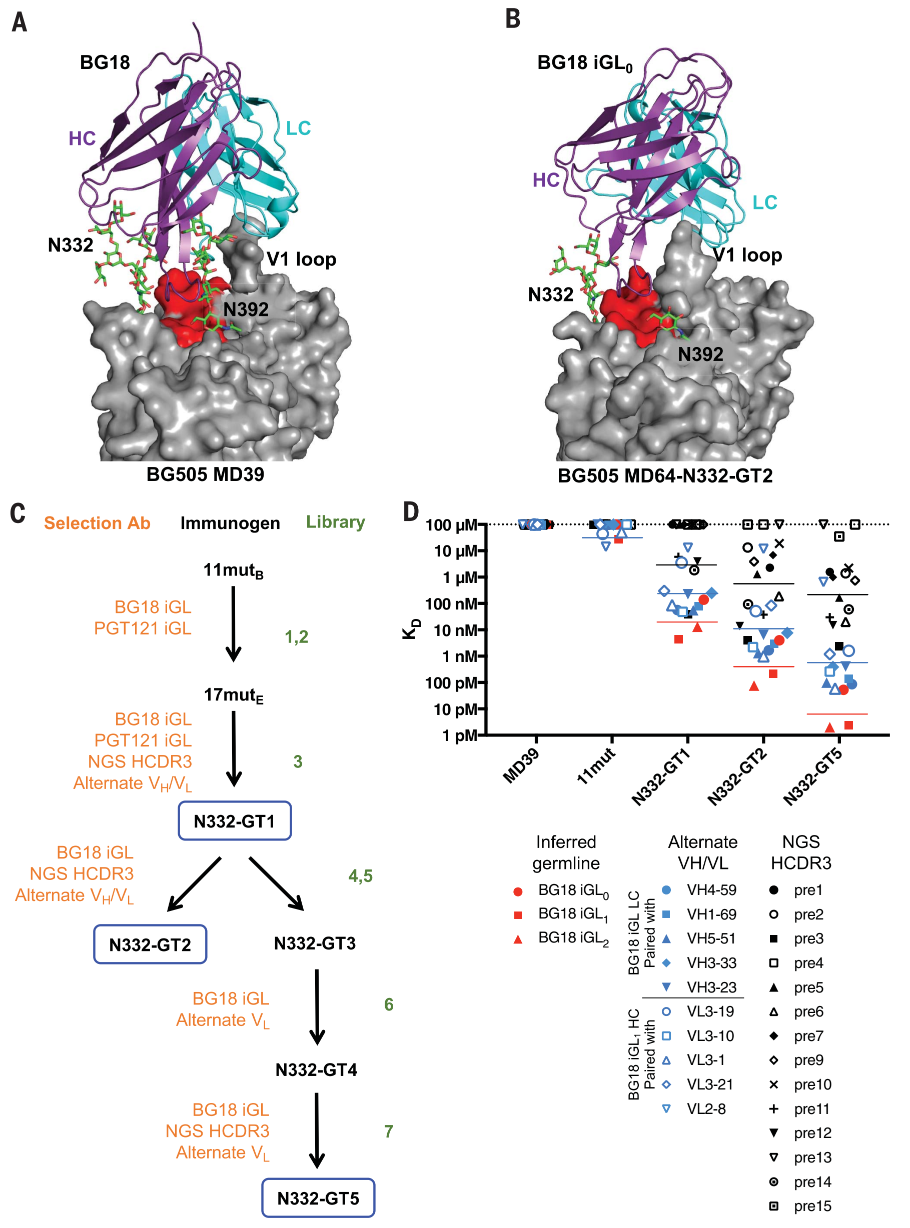


Fig. 1. Engineering germline-targeting trimers for an HCDR3-dependent bnAb. (**A**) Cryo-EM structure of BG18 (HC, purple; LC, cyan) bound to the BG505 MD39 Env trimer (gray, with N332 and N392 glycans shown as green sticks), and conserved residues near the base of V3 (Gly^{324}, Asp^{325}, Ile^{326}, Arg^{327}, Gln^{328}, Ala^{329}, His^{330}, Thr^{415}, Leu^{416}, and Pro^{417} colored red). (**B**) Cryo-EM structure of BG18 iGL$_0$ in complex with the N332-GT2 Env trimer with MD64-stabilizing mutations (*23*). The coloring is the same as in (A). (**C**) Schematic of the directed evolution process to design N332-GT1, -GT2, and -GT5. (**D**) N332-GT binding affinities (K_d) for BG18 iGL$_0$ to BG18 iGL$_2$ (red), BG18 iGL$_1$ with alternate germline V$_L$ (blue open symbols) or V$_H$ genes (blue filled symbols), and BG18 iGL containing NGS-derived HCDR3s (pre1 to pre15) (black). MD39 is the reference Env trimer with no germline-targeting mutations. Pre8 was found to be highly polyreactive and was not included in the analysis. Solid black, blue, and red lines indicate the geomean K_ds for NGS-derived precursors, alternate V$_H$ and V$_L$ precursors, and inferred germline precursors, respectively. The dashed line indicates the limit of detection.

vaccine targeting, namely that the high mutation level in BG18 itself is not required to achieve substantial neutralization breadth and potency in a BG18-like response. The elucidation of the BG18 binding mode by these studies enabled structure-guided immunogen design.

To assess the extent to which BG18-like precursor HCDR3s are present in the general population, we used a bioinformatics approach to search a custom next-generation sequencing (NGS) dataset of 1.1×10^9 sequences of human B cell receptor (BCR) heavy chains (HCs) from 14 healthy, HIV seronegative donors [8.58×10^8 sequences from four donors were obtained in this work (*20*) and were combined with 2.55×10^8 sequences from 10 donors from (*21*)]. Informed by our structural model for the BG18-Env interaction, we searched for BG18-like HCDR3 sequences with the same length as BG18, the same D gene in the same reading frame and position within the HCDR3, and the same heavy chain joining region (J$_H$) gene, allowing for diverse V-D and D-J junctions. Such BG18-like HCDR3 sequences were identified in all 14 donors (fig. S5), encouraging us to proceed with vaccine design. We further hypothesized that a range of BG18-like bnAbs utilizing alternate V$_H$ or V$_L$ genes could potentially interact with Env in a similar HCDR3-dependent binding mode. This hypothesis was subsequently supported by our ability to engineer BG18-like bnAbs utilizing three alternate V$_L$ genes (V$_L$3-1, V$_L$3-21, and V$_L$2-8) and two alternate V$_H$ genes (V$_H$4-59 and V$_H$4-61) (fig. S6). Identification of diverse BG18-like precursor HCDR3s from NGS data, and construction of BG18-like bnAbs with alternate V$_H$ or V$_L$ genes, led us to target a broad range of BG18-like precursors in the germline-targeting design process.

Design and antigenic analysis of immunogens

Germline-targeting immunogen design was carried out using a directed evolution method for engineering trimers on the surface of mammalian cells (*22*, *23*). We considered that it would be important to overcome the limitations of using only inferred-germline (iGL) antibodies (BG18 iGL$_0$ to iGL$_2$, fig. S5B) for the directed evolution of a germline-targeting immunogen with strong HCDR3 contacts. We reasoned that the germline-targeting design process directed to only iGL antibodies may fail to produce immunogens with appreciable affinity for diverse naïve precursors. iGL antibodies contain bnAb HCDR3 junctions that have been selected and most likely somatically mutated for high-affinity Env binding during bnAb affinity maturation. Therefore, such iGL antibodies may have features not present in the human antibody sequence repertoire. Furthermore, iGL antibodies likely underrepresent the diversity of potential precursors. We therefore designed a set of 15 BG18-like precursor antibodies that use BG18 germline-reverted genes but contain naïve human BG18-like HCDR3s with diverse junction regions identified in our search of NGS data described above (fig. S5B). On the basis of our finding that BG18-like bnAbs can utilize alternate V$_H$ and V$_L$ genes, we produced 10 additional BG18-like precursor antibodies with alternate V$_H$ or V$_L$ (fig. S7). This gave us 28 potential BG18-like precursors that could be used as selection reagents for directed evolution and multitarget optimization of Env trimer immunogens

capable of binding and priming a broad range of BG18-like precursor B cells.

Seven Env mammalian cell-surface display libraries, encoding amino acid variation within and around the BG18 epitope, were screened iteratively (20). At each stage, selection antibodies were used to isolate the highest affinity clones from the library, and the best mutations were incorporated into the next-generation Env immunogen. The first library was based on a previously described immunogen, 11mut$_B$ (22), that had weak but detectable affinity for BG18 iGL$_2$, the first selection antibody used (Fig. 1C, fig. S8, table S1, and supplementary text). In the early iterations, libraries were screened against the least challenging selection antibodies (e.g., BG18 iGL), whereas in later stages, the libraries were screened against more difficult antibody targets (e.g., NGS-derived and alternate V_H or V_L precursors) (Fig. 1C, figs. S5 and S7, and table S1). This directed evolution design process resulted in a series of germline-targeting Env trimers with increasing affinity for BG18 precursors (N332-GT1, -GT2, and -GT5; Fig. 1, C and D, fig. S8, and table S1). The N332-GT5 trimer bound with a dissociation constant (K_d) of ~2 pM to BG18 iGL$_1$, which represented a ~14 million–fold improvement over the initial protein design, 11mut$_B$. More importantly, whereas the 14 NGS-derived precursors tested had undetectable affinity to the initial protein design (and undetectable affinity for native HIV Env trimer MD39, Fig. 1D), the design process resulted in 11 of 14 acquiring affinity to N332-GT2 (geomean K_d = 519 nM, Fig. 1D) and 12 of 14 binding to N332-GT5 (geomean K_d = 234 nM, Fig. 1D). [One of the 15 NGS-derived precursors was found to be highly polyreactive and was therefore not included in our surface plasmon resonance (SPR) analyses]. Additionally, although only 3 of 10 alternate V_H or V_L precursor antibodies bound the starting protein design with low affinity (K_d > 10 μM) and none bound native HIV Env trimer (Fig. 1D), all 10 bound to N332-GT2 and N332-GT5 trimers, with robust affinities (geomean K_d = 11 nM and 572 pM, respectively, Fig. 1D). A K_d ≤ 1 μM may represent an affinity benchmark for generating robust germinal center (GC) responses from rare B cell precursors in the presence of polyclonal competitors in vivo (24), and 20 of 27 potential bnAb precursors bound to the N332-GT5 Env trimer with affinities of K_d ≤ 1 μM (Fig. 1D). Thus, the design process succeeded in extensively improving the immunogen binding properties to potential bnAb precursors with diverse HCDR3s and a variety of V_H and V_L genes.

Immunogen structural analysis

A cryo-EM–derived structure of BG18 iGL$_0$ complexed with the N332-GT2 trimer at ~3.9-Å resolution (Fig. 1B and table S2) showed that BG18 iGL$_0$ HCDR3 made a similar interaction to the base of the V3 as the HCDR3 of mature BG18 bound to the native-like trimer MD39 (Fig. 1A). Most of the additional interactions of BG18 iGL$_0$ complexed with N332-GT2 arise from V1 mutations in N332-GT2 that occupy a groove in the LC and also contact HCDR3 (figs. S9 and S10). HCDR3 dominates the interaction in the BG18 iGL$_0$ complex with N332-GT2, accounting for 64% of the total buried surface area. In the mature BG18 complex with the MD39 Env trimer, HCDR3 maintains the same key interactions and contributes 35% of the total buried area as the antibody makes substantially increased contacts to glycans N332, N392, and N137 (table S3). Overall, cryo-EM structures showed that N332-GT2 binds to BG18 iGL$_0$ with a similar HCDR3-dependent, V1-straddling binding mode as the BG505 MD39 Env trimer does with BG18.

Immunogenicity testing in a mouse model with rare bnAb precursors

To test the immunogenicity of the N332-GT2 Env trimer, we used a BG18^{gH} knock-in mouse engineered with a CRISPR-Cas9 rapid targeting strategy, in which ~30% of B cells express the BG18 iGL$_2$ HC variable region and mouse constant region paired with mouse LCs (25). The N332-GT2 Env trimer (but not MD39) bound to 12 ± 1% of naïve B cells in this mouse compared with 0.06 ± 0.01% in wild-type (WT) (C57BL/6) mice, demonstrating N332-GT2 specificity for BG18^{gH} naïve B cells (Fig. 2, A and B). Antigen-specific single–B cell sorting and BCR sequencing demonstrated that the N332-GT2–specific naïve BG18^{gH} B cells carry a variety of mouse LCs paired with BG18^{gH} (Fig. 2C). Furthermore, N332-GT2 had similar affinities for naïve BG18^{gH} B cell Fabs (geomean K_d of 582 nM) as for NGS-derived human BG18-like precursors (geomean K_d of 519 nM), showing the physiological relevance of the BG18-like precursor affinities in this mouse model.

To generate a mouse model with rare bnAb precursor B cells, we carried out adoptive transfer experiments in which 5000 CD45.2 BG18^{gH} B cells were transferred to CD45.1 WT mice on day –1, establishing a frequency of approximately seven GT2^{++}/KO$^-$ BG18^{gH} CD45.2 B cells per million CD45.1 B cells by day 0 (fig. S11) (KO indicates knockout). Control transfers were 50,000 CD45.2 WT B cells. Previously, we constructed ferritin nanoparticles (NPs) that displayed up to eight copies of MD39 native-like trimers (26), and mouse immunization studies showed that such NPs were superior to MD39 trimers in trafficking to follicular dendritic cell networks, concentrating in GCs, and eliciting immunoglobulin G (IgG) responses (27). We therefore engineered ferritin NPs displaying N332-GT2 trimers (fig. S12). Recipient mice were immunized at day 0 with either N332-GT2-NPs or control NPs displaying MD39 trimers lacking GT mutations, for a total of four immunization conditions (BG18^{gH} or WT B cells transferred, N332-GT2- or MD39-NPs immunized). Splenocytes were analyzed by flow cytometry at day 8 (Fig. 2, D and E, and fig. S13). GC B cells (CD38^{low}CD95$^+$) were detected in all four immunization conditions, but CD45.2 GC B cells were detected only in the case of N332-GT2-NP immunization of BG18^{gH} B cell recipients, demonstrating that N332-GT2-NPs activated rare BG18^{gH} B cells in vivo but MD39-NPs did not (Fig. 2D). N332-GT2-NPs induced CD45.2 GC B cells that bound to N332-GT2 but not to N332-GT2-KO (Fig. 2E) and were thus epitope-specific, consistent with a BG18-like response. By contrast, the same NPs induced considerably weaker epitope-specific responses among host CD45.1 GC B cells (Fig. 2E). In day 14 serum-binding analyses, N332-GT2-NPs induced strong epitope-specific IgG responses in BG18^{gH} B cell recipients and 15-fold weaker epitope-specific responses in WT B cell recipients (Fig. 2F), qualitatively consistent with the day 8 GC data. This demonstrated that activation of rare BG18^{gH} precursor B cells led to potent serum antibody responses and also showed that WT B cells responded to the BG18 epitope on N332-GT2. By contrast, MD39-NPs induced negligible BG18 epitope–specific serum responses in either BG18^{gH} or WT B cell recipients (Fig. 2F). Together, these results demonstrated that N332-GT2-NPs elicited GC and antibody responses from rare BG18^{gH} B cells.

By single-cell sorting and BCR sequencing CD45.2$^+$/N332-GT2^{++}/KO$^-$ GC B cells from BG18^{gH} recipient mice immunized with N332-GT2-NPs, we obtained HC-LC pairs at days 8 and 42. Of the HCs, 100% were derived from BG18^{gH}, formally proving that these GC responses utilized the knock-in HC (Fig. 2G). In contrast to the wide variety of mouse kappa genes used in LCs of N332-GT2–specific naïve BG18^{gH} B cells, by day 8 the LCs from GC B cells were highly enriched for two mouse kappa genes: *Igkv12-46* and *Igkv12-44* (Fig. 2G). By day 42, GC BCRs showed substantial somatic hypermutation, diversification, and affinity maturation compared with naïve B cells or day 8 GC BCRs (Fig. 2, H and I, and fig. S14). BG18^{gH} BCR Fab affinities for N332-GT2 trimers increased by a factor of ~6 from day 0 to day 8 (geomean K_ds of 582 and 97 nM, respectively; Fig. 2H). BG18^{gH} BCR Fab affinities increased dramatically by a factor of ~900 from day 0 to day 42 (geomean K_d = 640 pM, Fig. 2H). We conclude that N332-GT NPs can induce sustained GC responses and considerable affinity maturation and diversification from rare BG18-like precursors with human physiological affinities (see below), even in the presence of polyclonal competition.

To assess whether the affinity maturation induced by this single priming immunization was on a potential path toward bnAb

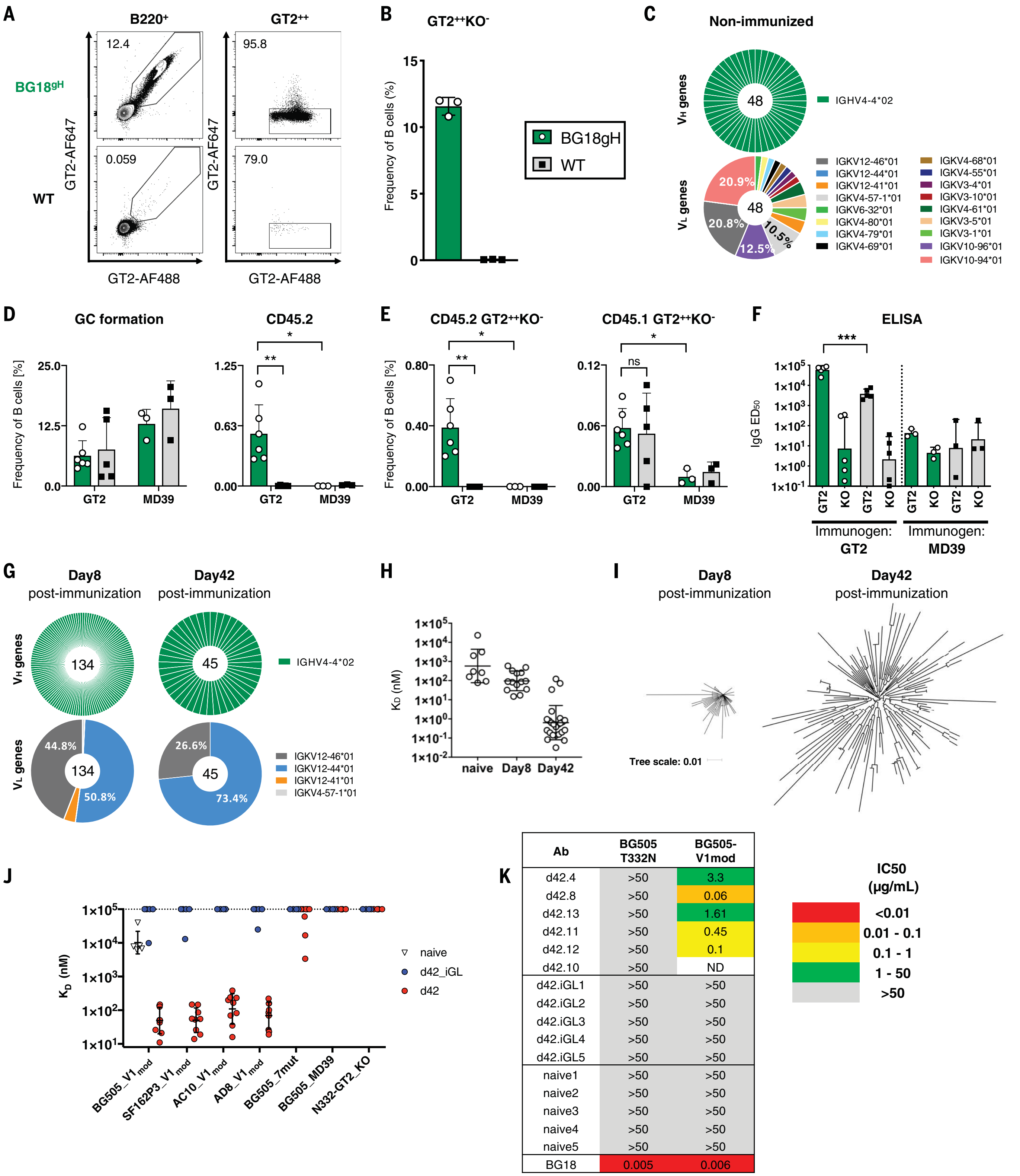
development, we tested whether day 42 antibodies could bind Env trimers more native-like than the germline-targeting immunogen or neutralize viruses with more native-like Env. The N332-GT2 immunogen has 17 germline-targeting mutations, eight of which are in two highly conserved regions of HIV Env (base of V3 loop around the GDIR motif and β19) and nine of which are in one highly variable region (V1 loop) (fig. S8). Thus, a key question was whether antibodies induced by N332-GT2 could recognize Env trimers with more native sequences lacking mutations in the two conserved regions. We therefore constructed a stabilized BG505 Env trimer that included eight of the N332-GT2 mutations in the V1 loop but was otherwise native-like (BG505-V1$_{mod}$) (fig. S15 and table S1), and we tested nine of the day 42 Fabs (those with highest affinity for N332-GT2) for their ability to bind this V1-modified Env trimer in SPR. All nine Fabs bound BG505-V1$_{mod}$, with a geomean K_d of 49 nM (Fig. 2J). By contrast, five naïve Fabs (those with highest affinity for N332-GT2) bound BG505-V1$_{mod}$ ~200-fold more weakly, with a geomean K_d of 10 μM, and inferred-germline variants of the day 42 Fabs (day 42 iGL Fabs) either showed no detectable affinity (four of five tested) or bound weakly (10 μM) (Fig. 2J). Furthermore, three of the day 42 Fabs, but none of the day 42 iGL Fabs, bound to the BG505 "7mut" trimer that is only six mutations away from a native-like Env trimer and was previously shown to be on the path toward development of PGT121-class bnAbs (22, 28) (Fig. 2J, fig. S15, and table S1). The day 42 antibodies did not bind N332-GT2-KO, consistent with BG18-like binding (Fig. 2J). None of the day 42 Fabs had detectable binding to the native-like trimer BG505 MD39, which was not surprising given the 17-mutation difference between the N332-GT2 immunogen and MD39 (Fig. 2J and fig. S8). Neutralization assays with BG505 WT and V1$_{mod}$ HIV pseudoviruses were consistent with our SPR findings: five of six day 42 Fabs tested neutralized V1$_{mod}$ HIV but not WT HIV, and none of the naïve or day 42 iGL Fabs neutralized either virus (Fig. 2K). We conclude that a single N332-GT2-NP priming immunization elicited functional BG18-like antibodies that could bind and neutralize viruses bearing Env that retains HIV-conserved regions and is more native-like than the immunogen.

To assess whether the affinity maturation due to priming conferred a degree of reactivity breadth beyond clade A BG505, we tested whether day 42 Fabs could bind to HIV Env trimers from three different isolates and two additional clades (SF162P3 and AC10, clade B; AD8, clade C), all with the same modified V1 loop as BG505-V1$_{mod}$ (fig. S15 and table S1). All nine day 42 Fabs tested bound to the three Env trimers with highly heterologous sequences, with geomean K_ds of 50, 110, and 69 nM for SF162-V1$_{mod}$, AC10-V1$_{mod}$, and AD8-V1$_{mod}$, respectively. By contrast, four of five day 42 iGL Fabs had no detectable affinity for these trimers (Fig. 2J). These data show that priming with N332-GT2 in this mouse model induced antibodies with a substantial degree of breadth in that they can bind with relatively high affinity to diverse stabilized Env trimers that share the same V1 loop.

Immunogen reactivity with naïve human B cells

A critical test of the germline-targeting design process was to determine if the N332-GT Env trimers could bind rare bnAb precursor human naïve B cells (29). To our knowledge, this is a human immunogen design benchmark that has only been met previously by the germline-targeting immunogen eOD-GT8 that targets VRC01-class bnAb precursors (9, 15). Attempts to isolate PGT121-related bnAb precursors using 11mut$_B$-related trimers did not succeed (supplementary text and fig. S16), consistent with our hypothesis that germline-targeting design using only iGL antibodies would be unsuccessful because of an inability to accommodate the natural sequence diversity among bnAb precursors in human B cell repertoires. To probe human naïve B cell reactivity to N332-GT Env trimers, we used N332-GT1 and N332-GT2 as sorting reagents and either BG505-MD39 Env (containing a native N332 epitope) or N332-GT2-KO Env (an epitope knockout) as negative sorting probes (Fig. 3A). About 16 million naïve B cells from six donors were probed with N332-GT1, and 62 million naïve B cells from 10 donors with N332-GT2, after accounting for polymerase chain reaction (PCR) and sorting efficiencies (table S4). All donors for ex vivo B cell sorting were distinct from the 14 NGS donors mentioned above (20). N332 glycan supersite epitope–specific naïve B cells [termed high-mannose patch clones (HMP) here] were isolated at a frequency of ~0.001% (Fig. 3B and fig. S17). These epitope-specific B cells were enriched for long HCDR3s (Fig. 3C). The B cells were also highly enriched for V_L3-25 and V_L3-1 LCs (Fig. 3D), which corresponded to the BG18 V_L and a V_L that we showed could be used by BG18-like precursors and bnAbs (Fig. 1D and fig. S6). We expressed and purified Fabs from 46 HMP naïve B cell clones (table S5) for

Fig. 2 Immunization of BG18^{gH} B cell adoptive transfer recipient mice with N332-GT2 Env NPs. (**A**) Gating strategy to identify epitope-specific (N332-GT2^{++}/N332-GT2-KO^{-}) B cells in BG18^{gH} and WT mice. (**B**) Frequency of epitope-specific B cells in nonimmunized BG18^{gH} and WT mice. Each symbol represents a different mouse. Bars indicate mean ± SD from experiments in three mice in each model. (**C**) Distribution of V_H and V_L genes in epitope-specific naïve B cells in nonimmunized BG18^{gH} mice. (**D**) Frequency of GC B cells (left) or CD45.2$^+$ GC B cells (right) in four immunization conditions. Each symbol represents a different mouse. Error bars indicate mean ± SD from experiments in the following number of mice in each condition: BG18^{gH} (GT2), n = 6; WT (GT2), n = 5; BG18gh (MD39), n = 3; and WT (MD39), n = 3. (**E**) Frequency of CD45.2$^+$ (left) or CD45.1$^+$ (right) epitope-specific B cells in four immunization conditions. Each symbol represents a different mouse. Error bars indicate mean ± SD from experiments in the following number of mice in each condition: BG18^{gH} (GT2), n = 6; WT (GT2), n = 5; BG18^{gH} (MD39), n = 3; and WT (MD39), n = 3. (**F**) Serum ELISA 50% equilibrium dilution (ED$_{50}$) values for N332-GT2 and N332-GT2-KO at day 14 after immunization for four immunization conditions. Each symbol represents a different mouse. Error bars indicate geometric mean and geometric SD from experiments in the following number of mice in each condition: BG18^{gH} (GT2), n = 5; WT (GT2), n = 5; BG18^{gH} (MD39), n = 3; and WT (MD39), n = 3. Student's t test was used. Not significant (ns) $P > 0.05$; *$P < 0.05$; **$P < 0.01$. Data in (A) to (F) are from one of three representative experiments with three or more animals in each group. (**G**) Distribution of V_H and V_L genes in epitope-specific GC (CD38^{low}CD95$^+$) B cells 8 and 42 days after immunization of BG18^{gH} B cell adoptive transfer recipient mice. (**H**) SPR dissociation constants for N332-GT2 trimer binding to epitope-specific Fabs derived from naïve B cells in nonimmunized BG18^{gH} mice and GC B cells 8 and 42 days after immunization of BG18^{gH} B cell adoptive transfer recipient mice. Each symbol corresponds to a different Fab and represents one or two measurements. Error bars indicate geometric mean and geometric SD. (**I**) Phylogenetic trees of BCR HCs isolated from epitope-specific CD45.2$^+$ B cells 8 and 42 days after immunization with N332-GT2 NPs. Tree scale indicates the number of substitutions per site. (**J**) SPR dissociation constants for the five highest affinity naïve Fabs from (H) binding to the V1 loop-modified BG505 trimer (BG505_V1$_{mod}$) and for nine of the high-affinity day 42 Fabs from (H) and five inferred-germline variants of the high-affinity day 42 Fabs (Day42.iGL) binding to V1 loop-modified trimers from BG505 and three other HIV isolates (SF162P3, AC10, and AD8), as well as a BG505 trimer with a less modified V1 loop (BG505_7mut), a native-like trimer (BG505_MD39), and an epitope-KO trimer (N332-GT2_KO). Each symbol corresponds to a different Fab and represents one or two measurements. Error bars indicate geometric mean and geometric SD. The dashed line indicates the limit of detection. (**K**) Neutralization potency (IC50) against native (BG505 T332N) and V1 loop-modified (BG505-V1mod) pseudoviruses for the BG18 bnAb, the five highest affinity naïve Fabs from (H), five inferred-germline variants of the high-affinity day 42 Fabs (d42.iGL), and five high-affinity day 42 Fabs (d42). Each IC50 is an average from two measurements. ND indicates not determined.



Ab	BG505 T332N	BG505-V1mod
d42.4	>50	3.3
d42.8	>50	0.06
d42.13	>50	1.61
d42.11	>50	0.45
d42.12	>50	0.1
d42.10	>50	ND
d42.iGL1	>50	>50
d42.iGL2	>50	>50
d42.iGL3	>50	>50
d42.iGL4	>50	>50
d42.iGL5	>50	>50
naive1	>50	>50
naive2	>50	>50
naive3	>50	>50
naive4	>50	>50
naive5	>50	>50
BG18	0.005	0.006

further examination of the biochemical properties and specificities of the naïve Abs. Twenty-three HMP Fabs bound to N332-GT1 and/or N332-GT2 by SPR and did not bind detectably to the MD39 native-like trimer, demonstrating proper N332 glycan supersite epitope specificity (Fig. 3, E and F). These SPR-validated epitope-specific clones were highly enriched for V_L3-25, V_L3-1, or the closely related V_L3-10 (Fig. 3E). Thus, the protein design strategy resulted in Env trimers that could successfully bind human naïve B cells with BG18-like LCs.

All N332-supersite bnAbs identified to date require a long HCDR3 [≥20 amino acids (aa)],

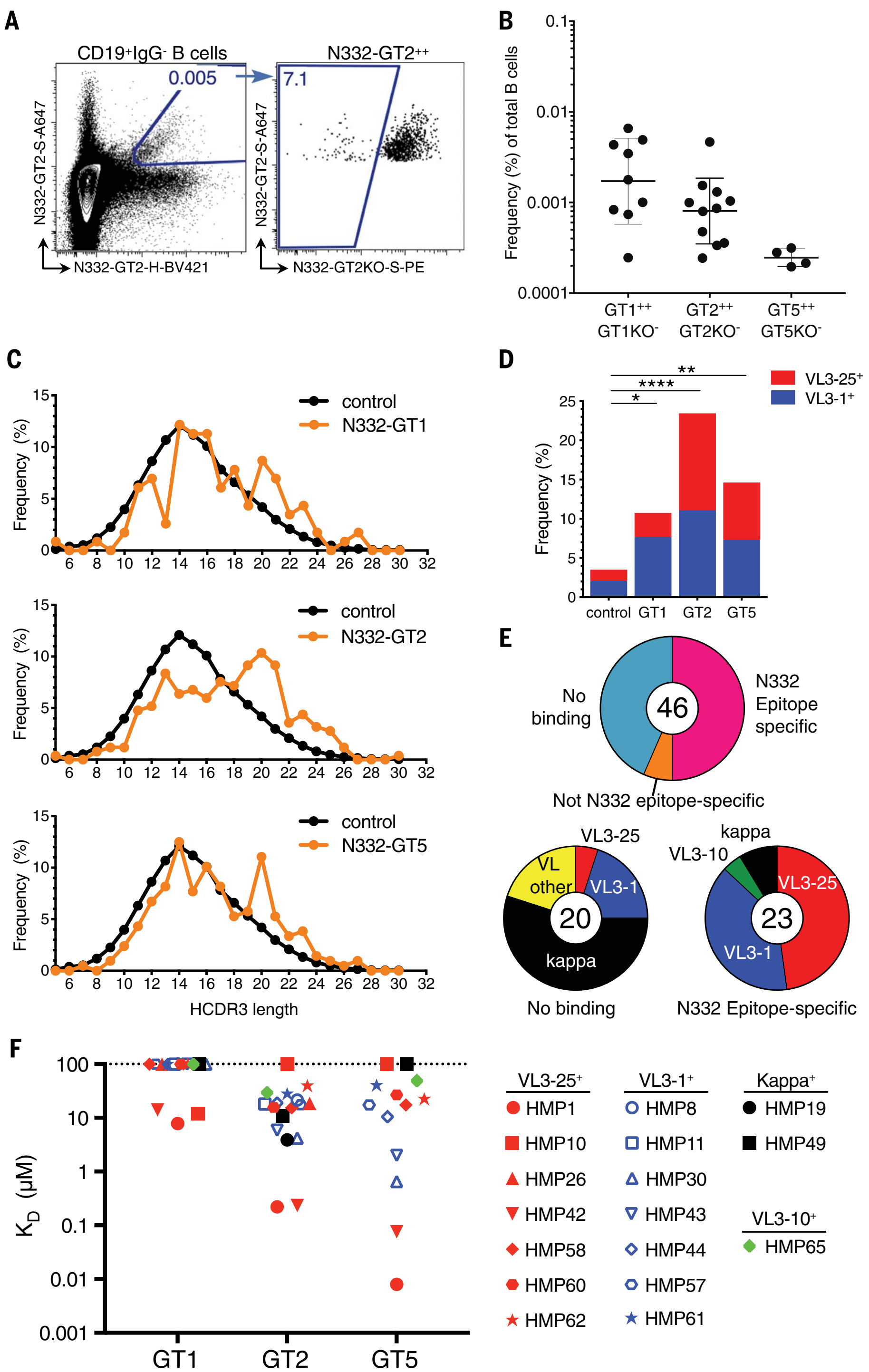


Fig. 3. Naïve human B cells sorted with N332-GT Env trimers. (**A**) Gating strategy for N332-GT epitope–specific sorting of naïve human B cells. (**B**) Frequency of epitope-specific B cells among IgG-negative B cells. Each symbol represents a different human subject. Error bars indicate geometric mean and geometric mean SD from the following number of independent subjects: N332-GT1, *n* = 9; N332-GT2, *n* = 11; and N332-GT5, *n* = 4. (**C**) HCDR3 length distribution from epitope-specific sorted cells compared with control B cells. (**D**) Frequency of V_L3-25 or V_L3-1 LCs from epitope-specific sorted cells relative to control B cells. Significance of differences from control was evaluated by a chi-square test. **P* = 0.01; ***P* = 0.005; *****P* = 0.0001. (**E**) SPR-derived binding specificities for 46 HMP Fabs corresponding to epitope-specific naïve human B cells isolated by N332-GT1 or N332-GT2 (top), with LC V gene usage for nonbinding Fabs (bottom left) and for N332 epitope–specific Fabs (bottom right). (**F**) SPR dissociation constants for HMP epitope–specific Fabs isolated with N332-GT1 and N332-GT2 Env trimers. The dashed line indicates the limit of detection.

owing to the structural requirements for the bnAb HCDR3 to reach the HIV Env protein surface at the base of V3 while avoiding V1 loop glycans (*19*). Although epitope-specific human B cells with HCDR3 lengths <20 aa were isolated using N332-GT1 or N332-GT2, only four of eight such clones tested by SPR were confirmed to be epitope specific, and their binding was weak (K_d >10 μM) (Fig. 3F and table S6). We considered that such B cells with HCDR3s <20 aa are probably unable to develop into N332-supersite bnAbs, and thus we did not study those clones further. Numerous epitope-specific naïve B cell clones with HCDR3s ≥20 aa were isolated with N332-GT1 and N332-GT2 probes (Fig. 4, A and B). From these human naïve B cell clones, we identified two categories of potential BG18-like precursors. The first category shared the same HCDR3 length, D gene, D gene reading frame, D gene position within HCDR3, and J_H gene with BG18 (Fig. 4A), exactly matching our initial search criteria when scanning NGS data for BG18-like HCDR3 sequences. Such naïve B cells were termed type I BG18-like precursors. The second category of epitope-specific BG18-like B cells had VL3-25, VL3-1, or VL3-10 and long HCDR3s (≥20 aa) with diverse HC sequences (Fig. 4B). We termed this more diverse class of isolated naïve B cells type II BG18-like precursors. HMP1 was a type I BG18-like precursor (Fig. 4A) with high affinity for the N332-GT2 Env trimer (K_d = 220 nM, Fig. 3F). The type II BG18-like precursors with confirmed binding exhibited a geomean K_d of 10 μM for the N332-GT2 Env trimer (Fig. 3F). Overall, type I and type II precursors accounted for 74% (17 of 23) of the HMP Fabs isolated by N332-GT1 or N332-GT2 and verified as epitope-specific by SPR (Fig. 3, E and F), indicating that such BG18-like precursors may represent a substantial fraction of the human naïve epitope-specific repertoire to these Env trimers. We then isolated additional type I and type II naïve B cell clones using N332-GT5 Env trimer probes with additional blood donors (Fig. 4, A and B). Overall, three type I BG18-like precursors were isolated at a frequency of ~1 in 53 million naïve B cells (HMP1, HMP68, and HMP69; table S4), in good agreement with our initial NGS bioinformatics-based estimate that precursors with BG18-like HCDR3s specific for N332-GT trimers may be present in the human B cell repertoire at a frequency of 1 in 54 million naïve B cells (fig. S5). Type II BG18-like precursors were isolated at a higher frequency of ~1 in 7 million naïve B cells, consistent with their larger sequence space.

Structural analysis of BG18-like human precursors bound to immunogens

To gain a structural understanding for the potential of human type I and type II BG18-like precursors (Fig. 4, A and B) to mature into

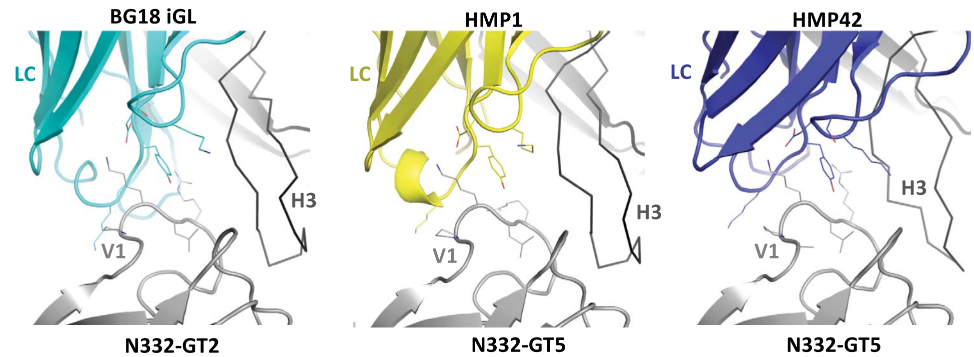
A

Name	Isolated with	HCDR3	Length	VH	D	JH	VL	JL
BG18-mat	-	ARNAIRIYGVVALGEWFHYGMDV	23	4-4	3-3	6	L3-25	2 or 3
BG18-iGL$_0$	-	ARNAITIFGVVIIGEYYYYGMDV	23	4-4	3-3	6	L3-25	2 or 3
HMP1	GT1	AREGFTIFGVVTFSEGYYYGMDV	23	4-61	3-3	6	L3-25	2 or 3
HMP68	GT5	TRAPITIFGVVQLGDYYYYYMDV	23	3-49	3-3	6	ND	ND
HMP69	GT5	ARDRITIFGVVIIGENYYYGMDV	23	3-11	3-3	6	L3-25	2 or 3

B

Name	Isolated with	HCDR3	Length	VH	D	JH	VL	JL
HMP30	GT2	AKDFLLSTSGYLSYGDYYYYGMDV	24	3-30	5-18	6	L3-1	2 or 3
HMP70	GT5	ARSWALFCSSTSCTNYYYYYGMDV	24	1-8	2-2	6	L3-25	3
HMP57	GT2	TTDIGYGDYLRSLDYYYYYGMDV	23	3-15	4-17	6	L3-1	2 or 3
HMP65	GT2	ARDQIREEWELLPEYYYYGMDV	22	3-30-3	1-26	6	L3-10	3
HMP71	GT5	AREESIVLMVYAIPADKDAFDI	22	1-2	2-8	3	L3-1	2 or 3
HMP72	GT5	ARKGPEGDGYNWELGTLYYFDY	22	4-59	5-24	4	L3-25	3
HMP8	GT1	ARQKVDYDFWSGYPAGYGMDV	21	3-74	3-3	6	L3-1	2 or 3
HMP26	GT2	ARDPTPNYYDSSGYWVGAFDI	21	4-59	3-22	3	L3-25	2 or 3
HMP42	GT2	ARDRGREWELESYYYYYYMDV	21	1-2	1-26	6	L3-25	2 or 3
HMP43	GT2	AKDIESRYFDWDNYYYYGMDV	21	3-9	3-9	6	L3-1	2 or 3
HMP44	GT2	ARDVSRNWGWEEDYYYYGMDV	21	3-30-3	2-21	6	L3-1	2 or 3
HMP58	GT2	ARVKYSSSWYVDDYYYGMDV	20	3-74	6-13	6	L3-25	2 or 3
HMP60	GT2	ARGDLYYYDSSGYYYYGMDV	20	3-30	3-22	6	L3-25	7
HMP61	GT2	ARAWDYDYVWGSYGVDGMDV	20	4-34	3-16	6	L3-1	2 or 3
HMP35	GT2	AKGPQHYYDSSGYPEYYFDY	20	3-23	3-22	4	L3-25	2 or 3
HMP38	GT1	ARDGTLYYYDSSGYYSYFDY	20	3-30-3	3-22	4	L3-1	2 or 3
HMP39	GT2	ARDRDSYYYDSSGYSDYFDY	20	3-33	3-22	4	L3-25	2 or 3
HMP40	GT2	ARGSVYYYDSSGYFWDAFDI	20	3-30-3	3-22	3	L3-1	1
HMP59	GT2	ARDATMSIAPNYYYYYGMDV	20	3-33	6-6	6	L3-25	2 or 3
HMP73	GT5	AREWDGGYSGYDSGDWYFDL	20	3-48	5-12	2	L3-1	2 or 3

C



D

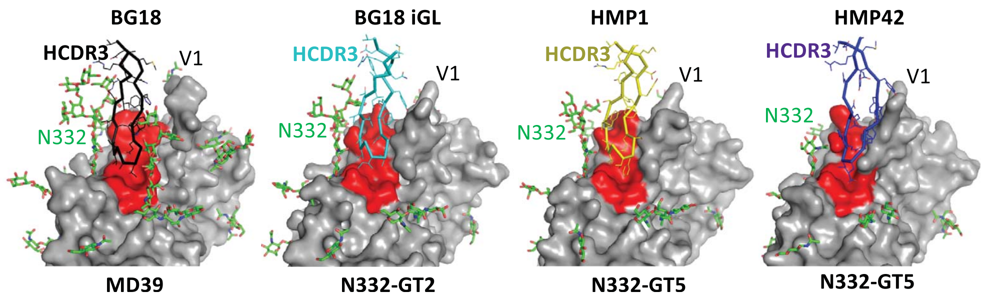


Fig. 4. Sequence and structural characterization of type I and type II BG18-like naïve antibodies isolated by N332-GT Env trimers. (**A**) HCDR3 sequence and gene segment assignments for three type I BG18-like naïve human precursor antibodies. V, D, and J gene segments are colored blue, red, and green, respectively. Single-letter abbreviations for amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**B**) HCDR3 sequence and gene segment assignments for 20 type II BG18-like precursor antibodies. (**C**) Cryo-EM structural analysis of type I (HMP1) and type II (HMP42) precursor antibody LC interactions with N332-GT5 compared with BG18 iGL LC interactions with N332-GT2. Gp120 is colored gray, and the LCs are colored cyan, yellow, and blue for BG18 iGL, HMP1, and HMP42, respectively. (**D**) Cryo-EM structural analysis of type I and type II precursor HCDR3 interactions with N332-GT5 compared with HCDR3 interactions for BG18+MD39 and BG18iGL+N332-GT2 complexes. Gp120 is colored gray with conserved residues (or the corresponding germline-targeting amino acids) near the base of V3 colored red as in Fig. 1A. Glycans are shown as green sticks.

bnAbs, we solved high-resolution cryo-EM structures of the complexes of N332-GT5 bound to HMP1 (type I) and HMP42 (type II), with resolutions of 3.7 and 3.4 Å, respectively (Fig. 4, C and D, and table S2). Both HMPs showed a similar LC binding mode as BG18 iGL, with the LC straddling the V1 loop (Fig. 4C and fig. S10). The HCDR3s of HMP1 and BG18 iGL have nearly identical conformations, supporting HMP1 and type I class naïve antibodies as ideal BG18 precursors (Fig. 4D). The projecting HCDR3 tip of HMP42 interacts with the same Env patch as BG18 but has a slightly different overall conformation that makes additional contacts with the Env V1 loop (Fig. 4D). This structural information supports the hypothesis that some or possibly all type II BG18-like naïve antibodies have a similar binding mode as BG18 iGL. Overall, these findings support the potential for both type I and type II BG18-like precursors to mature into BG18-like bnAbs (bnAbs with a BG18-like binding mode) under an appropriate sequential vaccination regimen. Given that type I and type II BG18-like precursors are enriched among epitope-specific human naïve B cells and have affinities that may confer competitive fitness in GCs, the data indicate that N332-GT Env trimers are strong candidates for priming BG18-like precursors for potential maturation into HIV bnAbs in humans.

Application to vaccine design for pathogens other than HIV

We explored whether our approach to target and prime a diverse pool of antibody precursors may have applicability to other pathogens. To evaluate whether our method of germline-targeting vaccine design could be applied beyond HIV, we carried out sequence and structural analyses for selected bnAb-antigen complexes for several major pathogens. In this nonexhaustive survey, we identified 11 potential antibody targets from five major pathogens, including hepatitis C virus (HCV), influenza virus, malaria, and dengue and Zika viruses (fig. S18). According to our sequence and buried surface area analyses, these antibodies all share the ability to make a series of important contacts with antigens through templated portions of their HCDR3s (portions encoded by D or J genes), which can be targeted by vaccine design. Most of the antibodies we identified are strongly HCDR3 dependent, on the basis of a criterion of HCDR3 contributing >30% of all surface area buried on the antibody. The strong HCDR3-dependence of the antibodies may allow for the development of related antibodies utilizing alternate V_H or V_L genes (as occurred with BG18) and hence may be advantageous for precursor frequency. With the exception of the dengue and Zika antibody EDE2 A11, all target antibodies have relatively common HCDR3 lengths of ≤22 aa represented by ≥2% of human antibodies (30), suggesting that HCDR3 length will not pose a limitation on precursor frequency. All target antibodies also have mutation levels in V_H and V_L that are present in ≥1 to 2% of human memory B cells (30), and all but two (9 of 11) lack indels, thus mutation level and indels should not pose a limitation on production of similar antibodies if appropriate precursors can be primed. In some cases, the native antigen has been shown to bind to an inferred-germline or unmutated common ancestor of the target antibody (31–33), raising the question of whether a germline-targeting approach would be necessary. We propose that even in such cases, our strategy may improve the design or validation of a vaccine priming candidate. Identification of a diverse set of antibody potential precursors with diverse HCDR3 junctions should allow for testing the breadth of precursor reactivity of the native antigen, and our design and validation strategies may optimize and/or verify breadth.

Concluding remarks

Most antibodies, and most HIV bnAbs, recognize their target in a strongly HCDR3-dependent manner. A central challenge of germline-targeting vaccine design is the large paratope sequence space and structural complexity possible for any set of antibodies targeting a conserved epitope by means of a shared HCDR3-dependent binding modality. Here, we demonstrate the successful design of a germline-targeting immunogen for this general class of antibody recognition. We used the human repertoire and structural features of bnAb-Env binding as guides to identify a pool of potential bnAb precursors and then design an immunogen with affinity for a representative set of those precursors. This procedure was validated by the isolation of three type I BG18-like precursors from naïve human B cells with N332-GT trimers and the demonstration that N332-GT NPs drove a robust BG18-class B cell response in an animal model with rare BG18 precursors. Furthermore, N332-GT trimer–sorted human naïve B cells were also enriched for type II BG18-like BCRs, and such precursors exhibited a BG18-like binding mode, indicating that the pool of potential BG18-like human naïve precursors is larger and more diverse than originally expected. This study does not demonstrate the induction of neutralizing antibodies to WT HIV isolates; the goal for germline-targeting priming immunogens is not to induce bnAbs directly but rather to induce bnAb-precursor B cell responses that have the potential to mature into bnAbs. Induction of bnAbs is the aim for a complete germline-targeting vaccine regimen, which would include a germline-targeting prime and a series of shepherding and polishing immunogens. Overall, we describe a new approach to define bnAb precursors for an epitope of interest and the design of vaccine priming immunogens that take advantage of that information. This approach lays out a generalizable pathway for the development and preclinical validation of germline-targeting immunogens for HCDR3-dependent antibody responses.

Materials and Methods

Experimental design

The overall objective of the study was to test a new, general method for design of germline-targeting immunogens to prime human naïve precursors to known bnAbs. Here we address the study design for the mouse model experiments.

Study objectives and experimental design

These experiments were designed primarily to test whether N332-GT NPs could generate GC responses with detectable levels of BG18^{gH} B cells in GCs, under preimmunization conditions of low BG18^{gH} naïve precursor B cell frequency and high polyclonal competition. BG18^{gH} naïve and GC B cells were identified by cytometry using cell surface markers, including the CD45.2 marker that distinguished these B cells from the WT host mouse B cells that were marked with CD45.1 Additionally, BG18^{gH} naïve and GC B cells were single-cell sorted using N332-GT and N332-GT-KO probes, and the epitope-specific BCRs were sequenced, in order to prove that the HCs were derived from the BG18 iGL$_2$ HC knock-in gene. This was not a foregone conclusion because the adoptively transferred B cells in these experiments were from a heterozygous knock-in with ~30% of B cells expressing the BG18 iGL$_2$ HC variable region. Finally, to assess the degree to which somatic hypermutation led to increased affinities in the BG18^{gH} B cells, soluble Fabs were expressed based on the sorted epitope-specific BCR sequences, and SPR studies were conducted to evaluate binding affinities to N332-GT immunogens. Additional corroborating information was gleaned from these studies by serum ELISA analysis.

Sample size

The number of mice in each group was limited by mouse availability and the costs and time associated with the experiments; however, the number of mice used was judged to be sufficient to detect clear differences between groups.

Randomization and blinding

Animal recipients of adoptive transfer were assigned to groups with no pattern. Neither randomization nor blinding were used, as they were not deemed necessary.

Data exclusion

No data were excluded.

Replicates

Data presented are from two independent experiments. The results have been reproduced in at least two additional experiments in this same mouse model and in a different but related mouse model.

BG18/BG505 SOSIP structural model

The HCDR3 loop of unliganded BG18 (PDB: 5UD9) was aligned to PGT122 (PDB: 4TVP) and several features suggested this as a plausible binding mode. First, $Arg^{L}54$ in LCDR2 would be positioned in a similar space as LCDR3 $Arg^{L}94$ in PGT122, a known critical contact residue for PGT122. Second, $Arg^{H}29$ was positioned close to the N332 glycan, and we confirmed that $Arg^{H}29$ was important for neutralization, by mutagenesis (fig. S2). Finally, analysis of computationally predicted V1 conformational ensembles including protein and glycan conformational diversity suggested that the LC could plausibly avoid clashing with the N137 glycan.

Design of minimally mutated versions of BG18

Design of minimally mutated versions of BG18 shown in fig. S4 was guided by analysis of the structural model of BG18 bound to BG505 SOSIP (fig. S2B), as this work was carried out before crystal structures were published of BG18 bound to BG505 and B41 SOSIP trimers (*18*) and before we obtained a cryo-EM structure of BG18 bound to the MD39 trimer (Fig. 1A). Framework and CDR mutations were reverted to germline if structural inspection indicated they were not contributing to the binding interaction. Several HC and LC variants were tested. BG18.11, which we refer to as "minBG18," was the least mutated variant that retained at least 50% of the breadth of BG18 while retaining similar potency as BG18. BG18.6 was the least mutated variant that showed any neutralization.

Design of BG18 bnAb variants using alternate V_H or V_L genes

The V_L variant engineering was carried out early in the study and was therefore guided only by our model for the BG18-Env interaction (fig. S2B). To engineer the V_L variants shown in fig. S6, the indicated V_L gene was substituted for the BG18 V_L gene and BG18 mutations were incorporated. The engineering of V_H variants was more complicated and was informed by our structural and mutagenesis studies (Fig. 1A and figs. S2, E and F, and S4). These studies indicated that the most important feature of the BG18 V_H gene (V_H4-4) was HCDR1 with a length of nine amino acids that is rare among human V_H genes. We therefore tested whether BG18-like bnAbs could utilize alternate V_H genes (4-59 and 4-61) that are closely related to V_H4-4 but use the more common HCDR1 lengths of 8 and 10 amino acids found among ~74 and ~19% of human V_H genes, respectively. We used mammalian display to screen scFv libraries containing ~10^4 to 10^5 HCDR1 sequences for binding to gp120 and native-like trimers based on several isolates (B41, 191084, ZM197, 6811) using the directed evolution design process described previously (*22*). One VH4-59 library included an NNK codon in the HCDR2 at position 53 in addition to containing HCDR1 sequence diversity. Briefly, libraries were integrated into 293T cells using a dox-inducible lentivirus based system; the scFv was anchored to the cell surface by linking the C-terminus to a PDGFR transmembrane domain; and the cells were incubated with HIS-tagged Env proteins and then stained with anti-HIS PE (miltenyi biotech). With this process, we identified two V_H4-59 clones and one V_H4-61 clone that when expressed as soluble IgG showed neutralization breadth on a BG18-sensitive virus panel (fig. S6). Because the theoretical number of HCDR1 sequences for lengths 8 and 10 are ~10^{10} and ~10^{13}, respectively, there are likely to be many HCDR1 sequences that can support neutralization beyond what we identified here. We conclude that BG18-like antibodies with diverse V_H and V_L genes can achieve broad and potent neutralization. It follows that BG18-like precursors containing alternate V_H and V_L genes should be targeted by vaccine design.

Immunogen design by mammalian cell surface display directed evolution

BG18 iGL_2 had detectable affinity to the $11mut_B$ (PGT121 germline targeting) trimer (*22*) but no detectable affinity to BG505 MD39 containing a native N332 epitope; therefore, we used $11mut_B$ as a base construct to begin the BG18 germline-targeting design process. The following libraries were screened using a previously described mammalian cell surface display method (*22*). Briefly, the following libraries were cloned into the pLenti CMVTRE3G Puro Dest plasmid and then stably integrated into rtTA3G-expressing HEK 293T cells using lentiviral transduction. Library 1 was a screen of all 20 aa at a subset of positions in the BG18/PGT121 epitope. It was an NNK codon scan of positions 294, 297, 298, 299, 300, 302, 304, 305, 326, 329, 330, 333, 386, 413, 414, 415, 416, 417, 419, and 420. NNK codons were introduced into BG505-$11mut_B$-gp120 using the QuikChange Site-Directed Mutagensis Kit (Agilent). Library 1 was screened for binding to BG18 iGL_2 and PGT121-$GL_{CDR3rev4}$. Library 2 was intended to sample hydrophobic amino acids underneath the BG18 epitope. It was a combinatorial library with amino acids F/I/L/V introduced at positions 154, 322, 323, 326, 333, 414, 415, and 416. The library insert was assembled with overlapping ultramers (IDT DNA) followed by Gibson cloning (NEB) into BG505-$11mut_B$-gp120. Library 2 was screened for binding to BG18 iGL_2 and PGT121-$GL_{CDR3rev4}$. Library 3 was intended to test all 20 aa at key epitope contact positions predicted by the structural model. It was a combinatorial library with NNK codons introduced at positions 137, 325, and (F/I/L/V) at position 326. The library insert was assembled with overlapping ultramers (IDT DNA) followed by Gibson cloning (NEB) into BG505-MD39-$17mut_E$. Library 3 was screened for binding to the following 24 Abs: BG18 iGL_0, BG18 iGL_1, pre1 - pre6, pre8, pre10 - pre15, V_L2-8, V_L2-14, V_L3-21, V_H1-69, V_H3-33, V_H4-59, V_H5-51, PGT121-$GL_{CDR3mat}$, and PGT121-$GL_{CDR3rev1}$. Library 4 was designed to test all 20 aa at key epitope contact positions predicted by the structural model and mutations at position 325 that were isolated in the library 3 screen. It was a combinatorial library with NNK codons introduced at positions 138 and 141 and (P/H/A/D) at position 325. The library insert was assembled with overlapping ultramers (IDT DNA) followed by Gibson cloning (NEB) into BG505-MD39-$17mut_E$-N137K. Library 4 was screened for binding to BG18 iGL_0, pre3, pre14, V_L3-21, V_L2-8, V_L2-14, V_H3-33. Library 5 was designed to test all amino acids at key epitope contact positions predicted by the structural model and mutations at position 325 that were isolated in the library 3 screen. It was a combinatorial library with NNK codons introduced at positions 138 and 139 and (P/H/A/D) at position 325. The library insert was assembled with overlapping ultramers (IDT DNA) followed by Gibson cloning (NEB) into BG505-MD39-$17mut_E$-N137K. Library 5 was screened for binding to BG18 iGL_0, V_L3-21, V_L2-8, V_L2-14, V_H1-69, and V_H5-51. Library 6 tested all 20 aa at positions not directly in the BG18 epitope to identify mutations that may indirectly effect binding to BG18 precursors. It was an NNK scan of positions 167 to 308. The insert was synthesized at SGI-DNA and Gibson cloned (NEB) into BG505-MD39-N332-GT3. Library 6 was screened for binding to BG18 iGL_0, V_L2-8. Library 7 screened all 20 aa at positions in and around the BG18 epitope, excluding the V1 loop. It was an NNK scan of positions 309 to 443. The insert was synthesized at SGI-DNA and Gibson cloned (NEB) into BG505-MD39-N332-GT3. Library 7 was screened for binding to BG18 iGL_0, pre1, pre2, pre4, pre10, pre15, and V_L2-8. All constructs contained a C-terminal myc tag and were anchored to the cell membrane via a C-terminal PDGFR transmembrane domain. Staining of the cell populations was typically done with IgG until saturated binding was obtained at low nanomolar IgG concentrations, and then Fabs were used for staining to maintain selection pressure. IgGs were labeled with Anti-Human IgG-R-PE (Sigma) and Fabs were labeled with Human IgG Fab PE (LSBio). Cell surface protein expression was detected using FITC conjugated chicken

anti-CMYC (ICL inc). Typically, libraries were sorted three to five times, and the enriched cell populations were frozen until sequencing could be carried out as described previously (22). Mutations found in the most enriched clones were incorporated into the most recent designs, synthesized at Genscript either as C-terminal His tagged gp120s or MD39/MD64 trimers in the pHLsec vector and expressed and purified as described previously (22).

Nanoparticle design and purification

To obtain multivalent immunogens, trimers were genetically fused to ferritin from *Helicobacter pylori* using a short flexible linker. Genes were codon optimized for HEK293 cells and cloned into the pHLsec plasmid (GenScript). MD39-NP DNA was cotransfected with a plasmid encoding human Furin protease into FreeStyle 293F cells (Invitrogen, Cat no. R79007) using 293Fectin (ThermoFisher) and proteins were expressed at 37°C for 4 days. NPs were purified either using snow drop lectin-conjugated agarose beads (Vector laboratories) or HiTrap NHS-Activated HP affinity columns (GE Healthcare) conjugated with PGT145, each followed by gel-filtration using a Superose 6 size-exclusion chromatography column (GE Healthcare). N332-GT2 NP had the Furin cleavage site replaced by a flexible linker (SHSGSGGSGSGGHA) as well as an L545P mutation, both discovered by library screening; hence, N332-GT2 NP was not cotransfected with Furin. NP-assembly was assessed by negative-stain EM and SEC + multiangle light scattering (SEC-MALS) using a Superose 6 10/300 column (GE Healthcare) at a flow rate of 0.5 ml/min followed by DAWN HELEOS II and Optilab T-rEX detectors (Wyatt Technology), correcting for the glycan molecular mass by applying the built-in protein-conjugate analysis (ASTRA).

Neutralization activity

Neutralizing activity of monoclonal antibodies (mAbs) was assessed using a single round of replication in TZM-bl target cells, in the absence of DEAE-dextran except for the assays in Fig. 2K, as described previously (34). Briefly, pseudoviruses were generated by cotransfection of HEK293T cells with an Env-expressing plasmid and an Env-deficient genomic backbone plasmid (pSG3ΔEnv).

Cryo-EM structure determination

High resolution cryo-EM structures were determined for four complexes: (i) MD39 + BG18; (ii) N332-GT2 + BG18iGL; (iii) N332-GT5 + HMP1; and (iv) N332-GT5 + HMP42. Of note, initial attempts to determine even low-resolution EM structures of HMP Fabs bound to N332-GT2 were not successful. However, in the late stages of this study, we found that the N332-GT5 trimer proved capable of forming stable complexes with the two clones, HMP1 and HMP42, representing BG18-type I and type II precursors, respectively. In general, trimers were incubated with a 6-10× molar excess of Fab overnight at room temperature. Complexes containing HMP1 or HMP42 also included RM20A3 Fab, a non-neutralizing trimer base-binding antibody that helps increase orientation sampling of the particles. The following morning, each complex was purified using a HiLoad 16/600 Superdex 200pg size-exclusion column (GE Healthcare) with Tris-buffered saline (50 mM Tris pH 7.4, 150 mM NaCl) as the running buffer, and the peak corresponding to trimer-Fab complex was pooled and concentrated to ~6 to 8 mg/ml. 3.5 μl of each complex was mixed with 0.5 μl of 0.42 mM n-dodecyl β-D-maltoside (DDM; Anatrace), such that final DDM concentration (0.06 mM) is below the critical micellar concentration (CMC). A 4-μl aliquot of the complex was applied to a C-Flat grid (CF-2/2-4C, Electron Microscopy Sciences, Protochips, Inc.) or Quantifoil grid (Q 1.2/1.3-4C, Quantifoil Micro Tools GmBH), which had been plasma cleaned for 10 s using a mixture of Ar/O_2 (Gatan Solarus 950 Plasma system), and following a 10-s incubation, the grid was blotted between 4 to 6 s and plunged into liquid ethane using an FEI Vitrobot Mark IV (100% relative humidity, 10°C).

The samples were imaged using either a Thermo Fisher Titan Krios operating at 300 kV or a Thermo Fisher Talos Arctica operating at 200 kV, both with a Gatan K2 Summit direct electron director operating in counting mode. Automated data collection was performed using the Leginon software suite (35). Each micrograph movie (250-ms exposure per frame) was collected at a magnification of 29,000× and a pixel size of 1.03 Å (Krios) or 36,000× and a pixel size of 1.15 Å (Arctica). Data collection statistics for each sample are summarized in table S1. Micrograph movie frames were aligned and dose-weighted using MotionCor2 (36), and CTF models were calculated using GCTF (37).

Single particles were selected using DoGPicker (38) from the whole-frame aligned and summed micrographs, and particles extracted using Relion 3.0 (39) using a box size of 288 or 320 pixels. After numerous rounds of 2D and 3D classification, final reconstructions were performed in Relion 3.0, and after postprocessing, the final resolution estimates (FSC 0.143) are ~3.9 Å for N332-GT2 + BG18iGL (C3 symmetry), ~4.4 Å for MD39 + BG18 (C3 symmetry), ~3.7 Å for N332-GT5 + HMP1 + RM20A3 (C3 symmetry), and ~3.4 Å for N332-GT5 + HMP42 + RM20A3 (asymmetric). Additional data processing statistics are summarized in fig. S3.

Atomic models were built and refined into the high-resolution reconstructions by creating homology models based off deposited coordinates of BG505 SOSIP.664 (PDB 5cez) and 354BG18 Fab (PDB 5ud9), as well as docking of an HMP42 Fab crystal structure from this study (table S6), followed by an iterative cycle of manual building in COOT (40) and real space refinement in Phenix 1.13 (41) and Rosetta Relax 3.10 (42). Glycans were validated by CARP (43) and Privateer (44), and overall structures were evaluated using EMRinger (45) and MolProbity (46). Buried surface area calculations were performed using UCSF Chimera (47).

ELISA

For analysis of serum responses from immunized mice

N332-GT2–specific antibody titers were detected by ELISA, using anti-His Ab (2 μg/ml) to capture N332-GT2 or N332-GT2-KO antigen (2 μg/ml) on the plate. Mouse sera were incubated for 2 hours and alkaline phosphatase conjugated anti-mouse IgG (Jackson ImmunoResearch, #115-055-071) was incubated another hour. Titers were determined from the dilution curve in the linear range of absorbance. All noncommercial ELISA plates were developed with p-Nitrophenyl Phosphate (Sigma, # N2770). Absorbance at 405 nm was determined with a plate reader (BioTek).

For analysis of mAb binding

mAb binding ELISAs were performed by capturing antigen (1 μg/ml) onto plates precoated with anti-His antibody (1 μg/ml; Genscript) and blocked with blocking buffer (5% skim milk, 1% fetal bovine serum, 0.2% tween 20 in PBS). Dilution series of mAbs were added as indicated and labeled with peroxidase-conjugated goat anti-mouse IgG (1:5000; Jackson ImmunoResearch). Wells were developed with 1-Step Ultra TMB-ELISA substrate (Thermo Scientific) diluted 1:4 in H_2O and stopped by addition of 0.5M H_2SO_4. Absorbance was read at 450 nm and reference absorbance measured at 570 nm was subtracted from each well.

Surface plasmon resonance (SPR)

Kinetics and affinities of antibody-antigen interactions were measured on a ProteOn XPR36 (Bio-Rad) using GLC Sensor Chip (Bio-Rad) or Biacore4000 (GE) with Series S Sensor Chip CM5 (GE). We used 1× HBS-EP+ pH 7.4 running buffer (20× stock from Teknova, Cat. No. H8022) supplemented with BSA at 1 mg/ml. Following manufacturer's instructions for Human Antibody Capture Kit (Cat. No. BR-1008-39 from GE), we immobilized about 6000 RUs of capture mAb onto each flow cell of GLC Sensor Chip or about 10,000 RUs in the case of the CM5 Sensor Chip. In a typical experiment on the ProteOn XPR36 system, about 300 to 400 RUs of mAbs were captured onto each flow cell, and analytes were passed over the flow cell at 50 μl/min for

3 min followed by a 5-min dissociation time. Regeneration was accomplished using 3 M magnesium chloride with a 180-s contact time and injected four times per cycle. Raw sensograms were analyzed using ProteOn Manager software (Bio-Rad), including interspot and column double referencing, and either Equilibrium fits or Kinetic fits with Langmuir model, or both, were used when applicable. For the Biacore4000 instrument, we used similar conditions but lower ligand capture levels. In the case of Fab-antigen kinetic and affinity measurements on ProteOn XPR36 or Biacore4000, we used a similar ligand-capture technique with several modifications. The capture reagent was His-tag Rabbit pAb (GenScript Cat. No. A00174). It was amine coupled to the Sensor Chip surface using the same protocol from the GE Human Antibody Capture Kit referenced above. Our regeneration solution was phosphoric acid 0.85% with a 30-s contact time, four injections per cycle. In the case of the ferritin nanoparticle experiment, we used the ProteOn XPR36 system and Human Antibody Capture protocol described above with one additional step. We captured PGT128 IgG at 1300 RU level in all channels, including reference, followed by NP (as ligand) capture at 1600 RU. All other steps were the same as in the Human Antibody Capture protocol. Analyte concentrations were measured on a NanoDrop 2000c Spectrophotometer using an absorption signal at 280 nm (*8*).

NGS dataset of human BCR HCs

This work utilized a large NGS dataset of 1.1×10^9 amino acid sequences of BCR HCs from 14 healthy, HIV-uninfected donors. In this dataset, 255 million sequences from 10 donors were obtained from (*21*), which used the HiSeq sequencing platform and an amplification strategy including unique identifiers (UIDs) to enable discrimination of unique mRNA transcripts from PCR artifacts. These 10 donors were evenly divided between males and females and nearly evenly divided between Caucasians and African Americans, and had ages ranging from 18 to 30 (*21*). These sequences were collapsed by UID, assigned to VDJ gene segments with Abstar (*21*), and then rendered unique by clustering at the 99% amino acid identity level within each of six biological replicates per donor. Thus the 255 million sequences were unique at the amino acid level within biological replicates. JSON output files from Abstar were converted to parquet format and uploaded to the Amazon S3 storage cloud. To query databases, Amazon Elastic Map Reduce (EMR) 5.15.5 was used to configure a Spark cluster with added PySpark and Zeppelin configurations. Zeppelin was used to assemble PySpark scripts to query the database with custom scripts. An additional 858 million sequences from four additional donors were obtained here by both HiSeq and NextSeq sequencing platforms without the use of UIDs, as described below.

BCR HC sequencing for four donors

Full leukopaks (three blood volumes) were obtained from four human subjects (AllCells LLC or Hemacare, Inc.) under a protocol approved by the Institutional Review Board of the respective commercial provider. All subjects were healthy, HIV-negative adults with no reported acute illness in the 14 days prior to leukapheresis, and samples were deidentified prior to shipment. The Institutional Review Board of The Scripps Research Institute determined that research with these samples did not constitute human subjects research. Immediately upon receipt of the leukopak, peripheral blood mononuclear cells (PBMCs) were purified by gradient centrifugation and cryopreserved in aliquots of approximately 5×10^8 PBMCs. The junctional regions of antibody heavy chain libraries were amplified as in Willis *et al.* (*48*). SPRI-purified sequencing libraries were initially quantified using fluorometry (Qubit, Thermo Fisher Scientific) before size determination using a bioanalyzer (Agilent 2100). Libraries were requantified using qPCR (KAPA Biosystems) before sequencing on either an Illumina HiSeq (2 × 150–bp chemistry) or NextSeq (2 × 150–bp chemistry). Sequences were merged with PANDAseq using the default (symple_bayesian) merging algorithm before annotation with Abstar (*21*). Identical amino acid sequences from the same donor and biological replicate were collapsed into a single unique amino acid sequence.

BG18 precursor frequency estimate

The NGS dataset of human BCR HCs was queried by bioinformatic searches to gain information on the frequency of BG18-like HCDR3s in the human B cell repertoire (fig. S5). HCDR3s meeting the definition of BG18-like feature set i in fig. S5A, constituting a broad set of potential BG18-like HCDR3 precursors, were identified in 14 of 14 donors (fig. S5C). The geomean frequency was 1 in 58,000 among the 10 donors sequenced by Briney *et al.* (*21*) using UIDs. To refine this frequency estimate, we considered that only 11 of 14 BG18 iGL variants with NGS-derived HCDR3s differing in the HCDR3 junctions (fig. S5B) exhibited binding to N332-GT2 (Fig. 1D). BG18-like feature set ii (HCDR3 junction features) in fig. S5A characterized amino acids present in the nontemplated junction regions of BG18 and its somatic variants (*17*) and in the 11 precursors that bound to N332-GT2. The frequency of HCDR3s meeting the definitions for both feature sets i and ii was found to be lower than those within set i by a factor of 104. Because the V_L gene plays an important role In the BG18 V1-loop straddling binding mode, we incorporated V_L gene usage into the frequency estimate, as feature set iii ("V_L gene family"). We made the conservative assumption that only V_L3 LCs can support the BG18-class binding mode, as all V_L3 LCs tested bound with high affinity to N332-GT2. The frequency of all V_L3-derived Abs in the HC-LC paired sequences in DeKosky *et al.* (*49*) was 1 in 9 (13845 V_L3s in 127701 sequences). We also assumed that any V_H gene can support this binding mode, because when five of the most common human V_H genes were substituted into BG18 iGL_1, all five variants showed low nanomolar binding to N332-GT2 (Fig. 1D). Therefore, no frequency factor was imposed for V_H gene usage. Multiplying the frequencies of all three feature sets together gave our best estimate for the frequency in the human B cell repertoire of BG18-like precursors that could be targeted by N332-GT2: 1 in 54 million.

N332-GT–specific naïve human B cell sorting and BCR sequencing

LRS (leukoreduction) tubes were obtained from the San Diego Blood Bank from healthy, HIV-seronegative human donors. These studies do not constitute human subjects research, as determined by the Institutional Review Boards of both La Jolla Institute and The Scripps Research Institute. More than 1 billion peripheral blood mononuclear cells were regularly recovered from each donor. $CD19^+$ B cells were isolated using a positive-selection magnetic-bead separation kit (Miltenyi Biotec) and resuspended in complete RPMI media with 10% FBS.

Avi-tagged protein immunogens were biotinylated using the Bulk BirA kit (Avidity, LLC). N332-GT5 and N332-GT5-KO probes were used in N332-GT5 sorting experiments. N332-GT2 and N332-GT2-KO probes were used in N332-GT2 sorting experiments. N332-GT1 and MD39 probes were used in N332-GT1 sorting experiments. $11mut_B$ and MD39 probes were used in $11mut_B$ sorting experiments. Biotinylated protein immunogens were individually premixed with fluorescently labeled streptavidin to form tetramer probes. Multiple tactics were used to avoid false positives: (i) used two "positive" probes, (ii) each "positive" probe used a different protein tag (His-tag or Strep-tag) to avoid tag specific B cells, (iii) used a "negative" probe to identify N332-epitope specific B cells, (iv) chose independent (no tandems) fluorochromes for all probes to avoid fluorochrome specific B cells (*29*). For example, N332-GT2 sorting experiments used the follow probes: N332-GT2–StrepTag-biotin + streptavidin Alexa Fluor 647 ("N332-GT2-S-A647"), N332-GT2–HisTag-biotin + streptavidin Brilliant Violet 421 ("N332-GT2-H-BV421"), and N332-GT2-KO–StrepTag-biotin + streptavidin phycoerythrin ("N332-GT2KO-S-PE").

Cells were incubated with N332-GT probes for 20 min at 4°C. Without washing, anti-CD19

(PE-Cy7,ThermoFisher, clone HIB19) and anti-CD20 (PE-Cy7, ThermoFisher, clone 2H7), in addition to anti-IgG (APC-Cy7, Biolegend, clone HP6017), anti-CD3 (APC eFluor780, ThermoFisher, clone UCHT1), anti-CD14 (APC eFluor780, ThermoFisher, clone 61D3), anti-CD16 (APC eFluor780, ThermoFisher, clone eBioCB16), and Live/Dead (APC eFluor780, ThermoFisher) for exclusion, were added for an additional 20 min. A BD FACSAria was used for all cell sorting. Cells were sorted at a flow rate of 1500 events/s using an 85-μm nozzle. Sorting stringency was set to a strict setting to obtain one cell per well. Single B cells were sorted directly into cold lysis buffer or N332-specific clonal B cell lines were generated and interrogated as in (*9*). cDNA synthesis, nested BCR PCR, Sanger sequencing and sequence analysis was carried out as in (*15*). Sorting was done with FACSDiva (BD) software and post-sort analyses were done with FlowJo (FlowJo, LLC).

In order to determine a normal HCDR3 length distribution for naïve human B cells, we combined sequences from DeKosky *et al.* (*49*) and Jardine *et al.* (*30*). For control V_L gene frequencies, only H+L paired sequences from DeKosky *et al.* (*49*) were used. Statistical analysis of LC V gene usage (Fig. 3D) was done by applying Fisher's exact test (two-sided) to each immunogen-probe B cell set (i.e., N332-GT1, N332-GT2, N332-GT5) compared to the reference normal population.

Adoptive transfer experiments and immunization

For adoptive transfer experiments, B cells were isolated from CD45.2 C57BL/6J ("WT") or BG18^{gH} KI mice of 8 to 10 weeks of age, and cells were resuspended in 150 μl of PBS and counted by The NucleoCounter NC-200 (ChemoMetec USA Inc). The 150-μl cell suspensions were injected i.v. into CD45.1 B6. SJL-*Ptprc*^{a} *Pepc*^{b}/BoyJ recipient animals (5 × 10^3 cells per mouse for BG18^{gH} transfers and 50 × 10^3 cells per mouse for C57BL/6 transfers). One day later, recipient mice were injected i.p. with 10 μg GT2- or MD39-NPs with Sigma adjuvant (Sigma, # S6322 SIGMA). After 8 days, mice were sacrificed to harvest spleen samples. Blood samples were taken from the submandibular vein on days 0 and 14 after immunization. Four immunization conditions were tested in two independent experiments, with the following total number of mice in each condition: (i) BG18^{gH} B cell transfer, N332-GT2 NP immunization (N = 6 for day 8 GC analysis, N = 5 for day 14 ELISA); (ii) BG18^{gH} B cell transfer, MD39 NP immunization (N = 3 for day 8 GC analysis, N = 3 for day 14 ELISA); (iii) WT B cell transfer, N332-GT2 NP immunization (N = 5 for day 8 GC analysis, N = 5 for day 14 ELISA); (iv) WT B cell transfer, MD39 NP immunization (N = 3 for day 8 GC analysis, N = 3 for day 14 ELISA). We complied with all relevant ethical regulations. The animal studies were approved by the Institutional Animal Care and Use Committee of Massachusetts General Hospital.

Antigen specific single-cell sorting for BCR sequencing in mouse experiments

Antigen tetramers were prepared by conjugating for 1 hour (room temp.) biotinylated N332-GT2 and N332-GT2-KO trimers with fluorescently labeled streptavidins (Alexa Fluor 488, Alexa Fluor 647, eBioscience; Alexa Fluor 568, Thermo Fisher Scientific) in a 4:1 molar ratio. The same streptavidins conjugated with biotinylated Fab anti-IgM and biotinylated BSA were used as positive and negative staining controls, respectively (data not shown). Single-cell suspensions generated from spleen samples were depleted of red blood cells by ACK lysis, Fc blocked (BD Biosciences), and stained in FACS buffer (2% FCS/PBS) with antigen tetramers for 30 min at 4°C, 50 nM concentration. Next, a cocktail of mAbs was added for 30 min at 4°C. For staining of splenocytes from naïve mice (Fig. 2A), the cocktail was B220 PerCP-Cy5.5 (Clone RA3-6B2, Biolegend), IgD PE-Cy7 (Clone 11-26c.2a, Biolegend), CD4 APC-eFluor780 (Clone RM4-5, eBioscience), CD8a APC-eFluor780 (Clone 53-6.7, eBioscience), F4/80 APC-eFluor780 (Clone BM8, eBioscience), Ly-6G APC-eFluor780 (Clone RB6-8C5, eBioscience). For staining of splenocytes from immunized mice (fig. S13), the cocktail was CD38 Alexa Fluor 700 (Clone 90, Invitrogen), CD45.2 PE (Clone 104, Biolegend), CD45.1 PerCP-Cy5.5 (Clone A20, Biolegend), B220 PB (Clone RA3-6B2, Biolegend), CD95 PE-Cy7 (Clone Jo2, BD Bioscience), CD4 APC-eFluor780 (Clone RM4-5, eBioscience), CD8a APC-eFluor780 (Clone 53-6.7, eBioscience), F4/80 APC-eFluor780 (Clone BM8, eBioscience), Ly-6G APC-eFluor780 (Clone RB6-8C5, eBioscience). Live-Dead staining kits (Thermo Scientific) were used to identify dead cells for exclusion from the analysis. Data acquisition and single-cell sorting were performed on FACS ARIA II (BD Bioscience) and analyzed with FlowJo v. 10 (Tree Star). Single-cell sorting and single-cell PCR were carried out as described previously (*25*).

Phylogenetic analysis

We used Clustal Omega to create a multiple sequence alignment (input amino acid sequences from heavy chain PCR) and iTol (EMBL) to plot phylogeny trees: https://itol.embl.de/help/gkw290.pdf. The distance between nodes in the trees in Fig. 2I is a representation of dissimilarity (or evolutionary distance); the scale indicates the number of substitutions per site.

Analysis of Ab-antigen complexes for pathogens other than HIV

A broad but nonexhaustive set of antibody-antigen complexes for diverse major human pathogens were analyzed. If the HCDR3 was judged to play an important structural role in the interaction, then the HCDR3 sequence was analyzed for similarity to human germline D genes, checking all reading frames for all D genes listed at IMGT (*50*). We did not have access to nucleotide sequences for all Abs, so amino acid sequences were used for this analysis. If a recognizable D gene could be identified, then the complex was subjected to buried surface area (BSA) analysis. Antibody BSA analysis was carried out using PDBePISA (*51*). Glycan interfaces were not included in the BSA analysis. All complexes analyzed for buried surface area were included in fig. S18, except if the HCDR3 length was >20 amino acids and the Ab BSA analysis indicated that V_H, V_L, and HCDR3 all contributed substantially to the buried area on the Ab, in which case the complex was not considered a promising target and was not considered further. CDRs and FWs were specified according to IMGT convention (*52*). Somatic hypermutation (SHM) was determined by aligning the V_H and V_L genes to the IMGT human germline V_H and V_L genes and calculating the % difference in amino acid sequence from the most similar human germline gene. Insertions or deletions relative to the most similar human germline V_H or V_L gene were identified similarly.

Statistical analysis

Statistical parameters including the mean (or geometric mean), the SEM (or geomeric standard deviation), and in some cases the P value, are reported in the figures. Statistical analyses were performed using Prism (GraphPad Software) and compared with Student's t test (Fig. 2, D to F) or chi-square test (Fig. 3D); P values < 0.05 were considered significant. Correspondences between the number of asterisks and the P values are stated in the figure legends.

REFERENCES AND NOTES

1. A. S. Fauci, An HIV vaccine is essential for ending the HIV/AIDS pandemic. *JAMA* **318**, 1535–1536 (2017). doi: 10.1001/jama.2017.13505; pmid: 29052689
2. S. A. Plotkin, Correlates of protection induced by vaccination. *Clin. Vaccine Immunol.* **17**, 1055–1065 (2010). doi: 10.1128/CVI.00131-10; pmid: 20463105
3. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337**, 183–186 (2012). doi: 10.1126/science.1225416; pmid: 22798606
4. R. Rappuoli, M. J. Bottomley, U. D'Oro, O. Finco, E. De Gregorio, Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *J. Exp. Med.* **213**, 469–481 (2016). doi: 10.1084/jem.20151960; pmid: 27022144
5. R. Andrabi, J. N. Bhiman, D. R. Burton, Strategies for a multi-stage neutralizing antibody-based HIV vaccine. *Curr. Opin. Immunol.* **53**, 143–151 (2018). doi: 10.1016/j.coi.2018.04.025; pmid: 29775847
6. J. Jardine *et al.*, Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340**, 711–716 (2013). doi: 10.1126/science.1234150; pmid: 23539181
7. A. T. McGuire *et al.*, Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. *J. Exp. Med.* **210**, 655–663 (2013). doi: 10.1084/jem.20122824; pmid: 23530120

8. J. G. Jardine et al., Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. Science 349, 156–161 (2015). doi: 10.1126/science.aac5894; pmid: 26089355
9. J. G. Jardine et al., HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. Science 351, 1458–1463 (2016). doi: 10.1126/science.aad9195; pmid: 27013733
10. A. T. McGuire et al., Specifically modified Env immunogens activate B-cell precursors of broadly neutralizing HIV-1 antibodies in transgenic mice. Nat. Commun. 7, 10618 (2016). doi: 10.1038/ncomms10618; pmid: 26907590
11. B. Briney et al., Tailored immunogens direct affinity maturation toward HIV neutralizing antibodies. Cell 166, 1459–1470.e11 (2016). doi: 10.1016/j.cell.2016.08.005; pmid: 27610570
12. M. Tian et al., Induction of HIV neutralizing antibody lineages in mice with diverse precursor repertoires. Cell 166, 1471–1484.e18 (2016). doi: 10.1016/j.cell.2016.07.029; pmid: 27610571
13. D. Sok et al., Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci transgenic mice. Science 353, 1557–1560 (2016). doi: 10.1126/science.aah3945; pmid: 27608668
14. M. Medina-Ramírez et al., Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo. J. Exp. Med. 214, 2573–2590 (2017). doi: 10.1084/jem.20161160; pmid: 28847869
15. C. Havenar-Daughton et al., The human naïve B cell repertoire contains distinct subclasses for a germline-targeting HIV-1 vaccine immunogen. Sci. Transl. Med. 10, eaat0381 (2018). doi: 10.1126/scitranslmed.aat0381; pmid: 29973404
16. U.S. National Library of Medicine, A phase I trial to evaluate the safety and immunogenicity of eOD-GT8 60mer vaccine, adjuvanted; https://clinicaltrials.gov/ct2/show/NCT03547245.
17. N. T. Freund et al., Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. Sci. Transl. Med. 9, eaal2144 (2017). doi: 10.1126/scitranslmed.aal2144; pmid: 28100831
18. C. O. Barnes et al., Structural characterization of a highly-potent V3-glycan broadly neutralizing antibody bound to natively-glycosylated HIV-1 envelope. Nat. Commun. 9, 1251 (2018). doi: 10.1038/s41467-018-03632-y; pmid: 29593217
19. D. Sok, D. R. Burton, Recent progress in broadly neutralizing antibodies to HIV. Nat. Immunol. 19, 1179–1188 (2018). doi: 10.1038/s41590-018-0235-7; pmid: 30333615
20. See materials and methods.
21. B. Briney, A. Inderbitzin, C. Joyce, D. R. Burton, Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature 566, 393–397 (2019). doi: 10.1038/s41586-019-0879-y; pmid: 30664748
22. J. M. Steichen et al., HIV vaccine design to target germline precursors of glycan-dependent broadly neutralizing antibodies. Immunity 45, 483–496 (2016). doi: 10.1016/j.immuni.2016.08.016; pmid: 27617678
23. D. W. Kulp et al., Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. Nat. Commun. 8, 1655 (2017). doi: 10.1038/s41467-017-01549-6; pmid: 29162799
24. R. K. Abbott et al., Precursor frequency and affinity determine B cell competitive fitness in germinal centers, tested with germline-targeting HIV vaccine immunogens. Immunity 48, 133–146.e6 (2018). doi: 10.1016/j.immuni.2017.11.023; pmid: 29287996
25. Y. C. Lin et al., One-step CRISPR/Cas9 method for the rapid generation of human antibody heavy chain knock-in mice. EMBO J. 37, e99243 (2018). doi: 10.15252/embj.201899243; pmid: 30087111
26. K. Sliepen et al., Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. Retrovirology 12, 82 (2015). doi: 10.1186/s12977-015-0210-4; pmid: 26410741
27. T. Tokatlian et al., Innate immune recognition of glycans targets HIV nanoparticle immunogens to germinal centers. Science 363, 649–654 (2019). doi: 10.1126/science.aat9120; pmid: 30573546
28. A. Escolano et al., Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice. Cell 166, 1445–1458.e12 (2016). doi: 10.1016/j.cell.2016.07.030; pmid: 27610569
29. C. Havenar-Daughton, R. K. Abbott, W. R. Schief, S. Crotty, When designing vaccines, consider the starting material: The human B cell repertoire. Curr. Opin. Immunol. 53, 209–216 (2018). doi: 10.1016/j.coi.2018.08.002; pmid: 30190230
30. J. G. Jardine et al., Minimally mutated HIV-1 broadly neutralizing antibodies to guide reductionist vaccine design. PLOS Pathog. 12, e1005815 (2016). doi: 10.1371/journal.ppat.1005815; pmid: 27560183
31. S. Bangaru et al., A site of vulnerability on the influenza virus hemagglutinin head domain trimer interface. Cell 177, 1136–1152.e18 (2019). doi: 10.1016/j.cell.2019.04.011; pmid: 31100268
32. D. Corti et al., A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. Science 333, 850–856 (2011). doi: 10.1126/science.1205669; pmid: 21798894
33. A. I. Flyak et al., HCV broadly neutralizing antibodies use a CDRH3 disulfide motif to recognize an E2 glycoprotein site that can be targeted for vaccine design. Cell Host Microbe 24, 703–716.e3 (2018). doi: 10.1016/j.chom.2018.10.009; pmid: 30439340
34. E. Landais et al., Broadly neutralizing antibody responses in a large longitudinal sub-Saharan HIV primary infection cohort. PLOS Pathog. 12, e1005369 (2016). doi: 10.1371/journal.ppat.1005369; pmid: 26766578
35. C. Suloway et al., Automated molecular microscopy: The new Leginon system. J. Struct. Biol. 151, 41–60 (2005). doi: 10.1016/j.jsb.2005.03.010; pmid: 15890530
36. S. Q. Zheng et al., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
37. K. Zhang, Gctf: Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
38. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 166, 205–213 (2009). doi: 10.1016/j.jsb.2009.01.004; pmid: 19374019
39. D. Kimanius, B. O. Forsberg, S. H. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. eLife 5, e18722 (2016). doi: 10.7554/eLife.18722; pmid: 27845625
40. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
41. P. D. Adams et al., PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
42. P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci. 23, 47–55 (2014). doi: 10.1002/pro.2389; pmid: 24265211
43. T. Lütteke, M. Frank, C. W. von der Lieth, Carbohydrate Structure Suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res. 33, D242–D246 (2005). doi: 10.1093/nar/gki013; pmid: 15608187
44. M. D. Winn et al., Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242 (2011). doi: 10.1107/S0907444910045749; pmid: 21460441
45. B. A. Barad et al., EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. Nat. Methods 12, 943–946 (2015). doi: 10.1038/nmeth.3541; pmid: 26280328
46. V. B. Chen et al., MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010). doi: 10.1107/S0907444909042073; pmid: 20057044
47. E. F. Pettersen et al., UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
48. J. R. Willis et al., Redesigned HIV antibodies exhibit enhanced neutralizing potency and breadth. J. Clin. Invest. 125, 2523–2531 (2015). doi: 10.1172/JCI80693; pmid: 25985274
49. B. J. DeKosky et al., In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat. Med. 21, 86–91 (2015). doi: 10.1038/nm.3743; pmid: 25501908
50. M. P. Lefranc et al., IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. 43, D413–D422 (2015). doi: 10.1093/nar/gku1056; pmid: 25378316
51. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (2007). doi: 10.1016/j.jmb.2007.05.022; pmid: 17681537
52. M. P. Lefranc et al., IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol. 27, 55–77 (2003). doi: 10.1016/S0145-305X(02)00039-3; pmid: 12477501

ACKNOWLEDGMENTS

We thank H. Gristick and P. Bjorkman for providing atomic coordinates of unliganded BG18 Fab in advance of publication (*17*) and C. Corbaci for graphical design assistance for the summary. **Funding:** This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) UM1 AI100663 (Scripps Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery) and UM1 AI144462 (Scripps Consortium for HIV/AIDS Vaccine Development) (to W.R.S., F.D.B., S.C., A.B.W., and D.R.B) and NIAID R01 AI113867 (to W.R.S.); by the Ragon Institute of MGH, MIT, and Harvard (to F.D.B., W.R.S., and D.R.B.); by the International AIDS Vaccine Initiative (IAVI) Neutralizing Antibody Consortium (NAC) and Center (to W.R.S., A.B.W., I.A.W., and D.R.B.); and through the Collaboration for AIDS Vaccine Discovery funding for the IAVI NAC Center (to W.R.S., A.B.W., I.A.W., and D.R.B.). **Author contributions:** J.M.S. and W.R.S. conceived the study. J.M.S., Y.-C.L., C.H.-D., S.P., G.O., D.R.B., A.B.W., S.C., F.D.B., and W.R.S. designed the study. J.M.S., D.W.K., S.R., A.R., and W.R.S. designed immunogens. J.M.S. designed Abs. Y.-C.L., S.P., S.K., E.M., and F.D.B performed immunization studies. C.H.-D. and L.T. performed naïve B cell–sorting studies. G.O. and J.L.T. performed cryo-EM studies. B.B. performed NGS. J.R.W. performed bioinformatics analyses. D.S., E.L., and J.U. performed neutralization assays. A.L., O.K., and X.H. characterized immunogens and Abs. E.G., N.P., Y.A., and M.K. purified proteins. S.M.B. and I.A.W. contributed structural information. J.M.S. and W.R.S. wrote the manuscript. All co-authors edited the manuscript. **Competing interests:** S.P. is now employed by GSK Vaccines S.r.l., a company that might benefit indirectly from this research. D.R.B. is a paid consultant of IAVI. J.M.S. and W.R.S. are inventors on a patent application submitted by IAVI and The Scripps Research Institute that covers the N332-GT immunogens developed in this manuscript. **Data and materials availability:** Coordinates and maps for the structural data presented in this manuscript have been deposited to the Protein Data Bank under accession codes 6DFG, 6DFH, 6NF5, 6NFC, and 6OC7 and to the Electron Microscopy Data Bank under accession codes EMD-7875, EMD-7876, EMD-7884, and EMD-7885. Antibody sequences discovered during this study have been deposited to GenBank under accession numbers MN495018 to MN495471 (BG18^{gH} mouse antibodies) and MN514889 to MN514945 (human naïve B cell antibodies binding N332-GT immunogens). Custom scripts for the NGS database query will be made available from W.R.S upon request. NGS sequencing data used in this manuscript and example analysis methods are available at https://github.com/SchiefLab/SteichenScience2019. All other data are available in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaax4380/suppl/DC1
Supplementary Text
Figs. S1 to S18
Tables S1 to S7
References (*53*–*62*)

View/request a protocol for this paper from *Bio-protocol*.

22 March 2019; accepted 17 October 2019
Published online 31 October 2019
10.1126/science.aax4380

RESEARCH ARTICLE SUMMARY

ASTEROIDS

Episodes of particle ejection from the surface of the active asteroid (101955) Bennu

D. S. Lauretta*† and C. W. Hergenrother*† *et al.*

INTRODUCTION: Active asteroids are small bodies in the Solar System that show ongoing mass loss, such as the ejection of dust, which may be caused by large impacts, volatile release, or rotational acceleration. Studying them informs our understanding of the evolution and destruction of asteroids and the origin of volatile materials such as water on Earth.

The OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft has rendezvoused with the near-Earth asteroid (101955) Bennu. The selection of Bennu as the OSIRIS-REx target was partially based on its spectral similarity to some active asteroids. Observations designed to detect mass loss at Bennu were conducted from Earth and during the spacecraft's approach, but no signs of asteroid activity were found. However, when the spacecraft entered orbit in January 2019, we serendipitously observed particles in the vicinity of Bennu that had apparently been ejected from its surface.

RATIONALE: We analyzed the properties and behavior of particles ejected from Bennu to determine the possible mechanisms of ejection and provide understanding of the broader population of active asteroids. Images obtained by the spacecraft indicate multiple discrete ejection events with a range of energies and resultant particle trajectories. We characterized three large ejection events that respectively occurred on 6 January, 19 January, and 11 February 2019. Tracking of individual particles across multiple images by means of optical navigation techniques provided the initial conditions for orbit determination modeling. By combining these approaches, we estimated the locations and times of ejection events and determined initial velocity vectors of particles. We estimated the particle sizes and the minimum energies of the ejection events using a particle albedo and density consistent with observations of Bennu.

RESULTS: Particles with diameters from <1 to ~10 cm were ejected from Bennu at speeds ranging from ~0.05 to >3 m s^{-1}. Estimated energies ranged from 270 mJ for the 6 January event to 8 mJ for the 11 February event. The three events arose from widely separated sites, which do not show any obvious geological distinction from the rest of Bennu's surface. However, these events all occurred in the late afternoon, between about 15:00 and 18:00 local solar time.

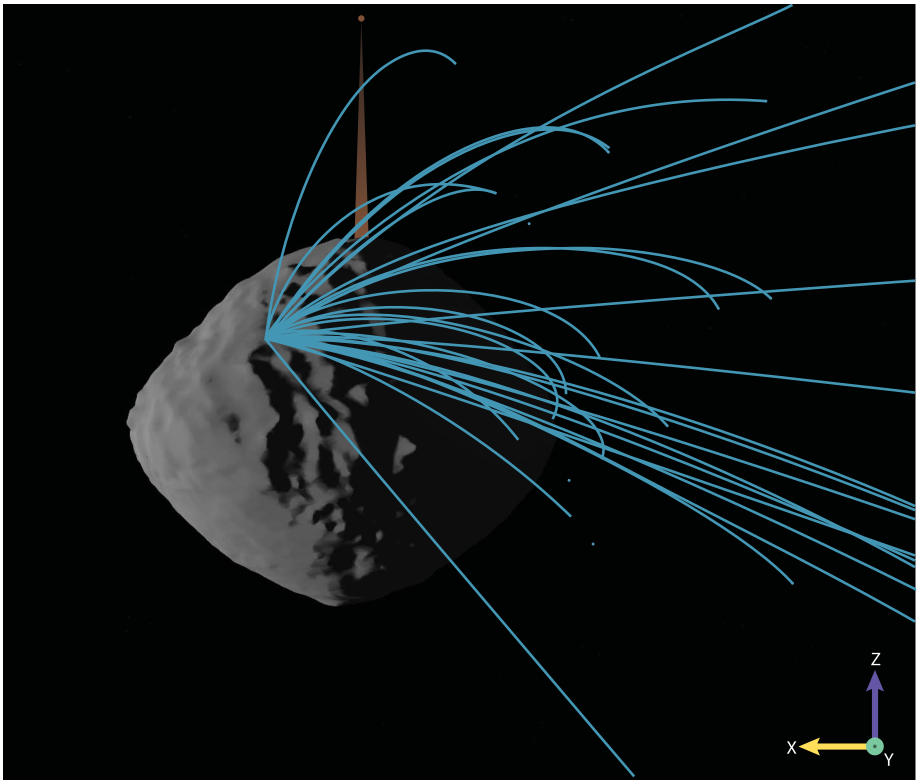
In addition to discrete ejection events, we detected a persistent background of particles in the Bennu environment. Some of these background particles have been observed to persist on temporary orbits that last several days—in one case, with a semimajor axis >1 km. The orbital characteristics of these gravitationally bound objects make it possible to determine the ratio of their cross-sectional area to their mass. Combined with their photometric phase functions, this information constrains the parameter space of the particles' diameters, densities, and albedos.

CONCLUSION: Plausible mechanisms for the large ejection events include thermal fracturing, volatile release through dehydration of phyllosilicates, and meteoroid impacts. The late-afternoon timing of the events is consistent with any of these mechanisms. Bennu's boulder geology indicates that thermal fracturing, perhaps enhanced by volatile release, could occur on the asteroid surface. Smaller events, especially those that occur on the night side of Bennu, could be attributable to reimpacting particles.

Our observations classify Bennu as an active asteroid. Active asteroids are commonly identified by major mass loss events observable with telescopes, on scales much greater than we observed at Bennu. Our findings indicate that there is a continuum of mass loss event magnitudes among active asteroids. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay3544



Schematic diagram of orbit determination model output for the 19 January 2019 particle ejection event from asteroid Bennu observed by the OSIRIS-REx spacecraft. Bennu is depicted in gray and has a diameter of ~500 m. OSIRIS-REx is indicated with the brown dot, ~2 km from Bennu's center of mass; the cone represents the viewing angle. Blue arcs are particle trajectories, ending or with gaps where the trajectories pass into shadow. The Sun–angular momentum frame coordinates are shown at bottom right: *x*, solar vector; *y*, Bennu orbital direction; *z*, Bennu north.

The full lists of author names and affiliations are available in the full article online.
*Corresponding author. Email: lauretta@orex.lpl.arizona.edu (D.S.L.); chergen@lpl.arizona.edu (C.W.H.)
†These authors contributed equally to this work.
Cite this article as D. S. Lauretta and C. W. Hergenrother *et al.*, *Science* 366, eaay3544 (2019). DOI: 10.1126/science.aay3544.

RESEARCH ARTICLE

ASTEROIDS

Episodes of particle ejection from the surface of the active asteroid (101955) Bennu

D. S. Lauretta[1]*†, C. W. Hergenrother[1]*†, S. R. Chesley[2], J. M. Leonard[3], J. Y. Pelgrift[3], C. D. Adam[3], M. Al Asad[4], P. G. Antreasian[3], R.-L. Ballouz[1], K. J. Becker[1], C. A. Bennett[1], B. J. Bos[5], W. F. Bottke[6], M. Brozović[2], H. Campins[7], H. C. Connolly Jr.[8,1], M. G. Daly[9], A. B. Davis[10], J. de León[11], D. N. DellaGiustina[1,12], C. Y. Drouet d'Aubigny[1], J. P. Dworkin[5], J. P. Emery[13,14], D. Farnocchia[2], D. P. Glavin[5], D. R. Golish[1], C. M. Hartzell[15], R. A. Jacobson[2], E. R. Jawin[16], P. Jenniskens[17], J. N. Kidd Jr.[1], E. J. Lessac-Chenen[3], J.-Y. Li[18], G. Libourel[19], J. Licandro[11], A. J. Liounis[5], C. K. Maleszewski[1], C. Manzoni[20], B. May[20], L. K. McCarthy[3], J. W. McMahon[10], P. Michel[19], J. L. Molaro[18], M. C. Moreau[5], D. S. Nelson[3], W. M. Owen Jr.[2], B. Rizk[1], H. L. Roper[1], B. Rozitis[21], E. M. Sahr[3], D. J. Scheeres[10], J. A. Seabrook[9], S. H. Selznick[1], Y. Takahashi[2], F. Thuillet[19], P. Tricarico[18], D. Vokrouhlický[22], C. W. V. Wolner[1]

Active asteroids are those that show evidence of ongoing mass loss. We report repeated instances of particle ejection from the surface of (101955) Bennu, demonstrating that it is an active asteroid. The ejection events were imaged by the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft. For the three largest observed events, we estimated the ejected particle velocities and sizes, event times, source regions, and energies. We also determined the trajectories and photometric properties of several gravitationally bound particles that orbited temporarily in the Bennu environment. We consider multiple hypotheses for the mechanisms that lead to particle ejection for the largest events, including rotational disruption, electrostatic lofting, ice sublimation, phyllosilicate dehydration, meteoroid impacts, thermal stress fracturing, and secondary impacts.

Active asteroids are small bodies that have typical asteroidal orbits but show some level of mass-loss activity, such as ejection of dust or the development of a coma or tail (*1*). Several objects in the main asteroid belt or the near-Earth asteroid population have been observed to show varying levels of mass loss, such as the active asteroid 133P/Elst-Pizarro (*2*). Some of these objects behave as comets and eject dust over long periods of time, from days to months, or during multiple perihelion passages [including 133P/Elst-Pizarro (*3*)]. Other active asteroids eject dust over short time scales in one or a series of impulsive events, such as in the case of (6478) Gault (*4*). Still others have been observed to split into multiple objects or, in the case of P/2016 G1 (PANSTARRS), completely disintegrate (*5*). Near-Earth asteroid (3200) Phaethon has exhibited low levels of mass loss during multiple orbits when less than 0.15 astronomical units (AU) from the Sun (*6*, *7*). Multiple ejection mechanisms have been suggested to explain asteroid activity, including collisions, water-ice sublimation, rotational destabilization, thermal fracturing, and dehydration (*8*).

The OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft arrived at the ~500-m-diameter B-type near-Earth asteroid (101955) Bennu in December 2018. Bennu was selected as the mission target partly because of its spectral similarity to some active asteroids (*9*). Here, we describe and analyze OSIRIS-REx observations of activity originating from Bennu's surface. We initially detected this phenomenon in navigational images from 6 January 2019, 1 week after the spacecraft entered orbit and 4 days before Bennu perihelion (*10*). We subsequently detected multiple particle ejection events between December 2018 and February 2019. The largest observed events each released dozens of observed particles.

[1]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [3]KinetX Aerospace, Simi Valley, CA, USA. [4]Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada. [5]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [6]Southwest Research Institute, Boulder, CO, USA. [7]Department of Physics, University of Central Florida, Orlando, FL, USA. [8]Department of Geology, Rowan University, Glassboro, NJ, USA. [9]The Centre for Research in Earth and Space Science, York University, Toronto, ON, Canada. [10]Smead Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO, USA. [11]Instituto de Astrofísica de Canarias and Departamento de Astrofísica, Universidad de La Laguna, Tenerife, Spain. [12]Department of Geosciences, University of Arizona, Tucson, AZ, USA. [13]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA. [14]Department of Astronomy and Planetary Sciences, Northern Arizona University, Flagstaff, AZ, USA. [15]Department of Aerospace Engineering, University of Maryland, College Park, MD, USA. [16]Smithsonian Institution National Museum of Natural History, Washington, DC, USA. [17]SETI (Search for Extraterrestrial Intelligence) Institute, Mountain View, CA, USA. [18]Planetary Science Institute, Tucson, AZ, USA. [19]Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS (Centre national de la recherche scientifique), Laboratoire Lagrange, Nice, France. [20]London Stereoscopic Company, London, UK. [21]School of Physical Sciences, Open University, Milton Keynes, UK. [22]Institute of Astronomy, Charles University, Prague, Czech Republic.
***Corresponding author. Email: lauretta@orex.lpl.arizona.edu (D.S.L.); chergen@lpl.arizona.edu (C.W.H.)** †These authors contributed equally to this work.

Particle detections

Dust and natural satellite searches were conducted during the spacecraft's approach to Bennu during early proximity operations in September to November 2018, which yielded null results (*10*). Signs of asteroid activity may have been detected by the OSIRIS-REx Laser Altimeter [OLA (*11*)] (figs. S1 and S2) in December 2018. OLA recorded 21 lidar returns off the limb of the asteroid during the Preliminary Survey mission phase, including four at distances of 399 m (4 December), 397 m (8 December), and 562 and 576 m (12 December, 3.1 hours apart) from Bennu's center. These four signals prompted a search for corresponding objects in images from the same dates, without success. However, the geometry suggests that these four returns were probably from objects or groups of objects (*12*). The earliest evidence of activity in imaging data is a particle 8 ± 3 cm (1σ) in diameter on a suborbital trajectory, imaged by the NavCam 1 imager of the Touch and Go Camera System (TAGCAMS) (*13*) on 10 December 2018. We cannot rule out activity before December 2018. The searches performed during the spacecraft's approach to the asteroid did not have sufficient sensitivity to detect most of the activity that was later observed at closer ranges. A particle as large as the one observed on 10 December would have been detectable with the natural satellite searches; lack of detection implies that events ejecting particles of that size were relatively rare or nonexistent during the spacecraft's approach.

On 31 December 2018, the spacecraft entered into an eccentric, near-terminator orbit that ranged between 1.6 and 2.1 km from Bennu's center of mass. This Orbital A mission phase continued until 28 February 2019, when the spacecraft departed orbit to perform the Detailed Survey (*14*). During the early part of Orbital A, we acquired NavCam 1 image sets roughly every 2 hours to provide optical navigation (OpNav) data for the flight dynamics team (table S1) (*15*). Each image set consisted of four images taken in pairs ~7 min apart. Each pair contained a short-exposure image (1.4 ms) to capture landmarks on Bennu's surface, followed immediately by a long-exposure image (5 s) to capture the background star field.

The first particle ejection event that we identified was observed in OpNav images taken on 6 January at 20:56:21 Coordinated Universal Time (UTC) (Fig. 1, A and B, and fig. S3). The particles appear as more than 200 starlike point-source objects and trailed (higher-velocity) objects located off the northern polar limb of Bennu. The image taken 7 min and 16 s

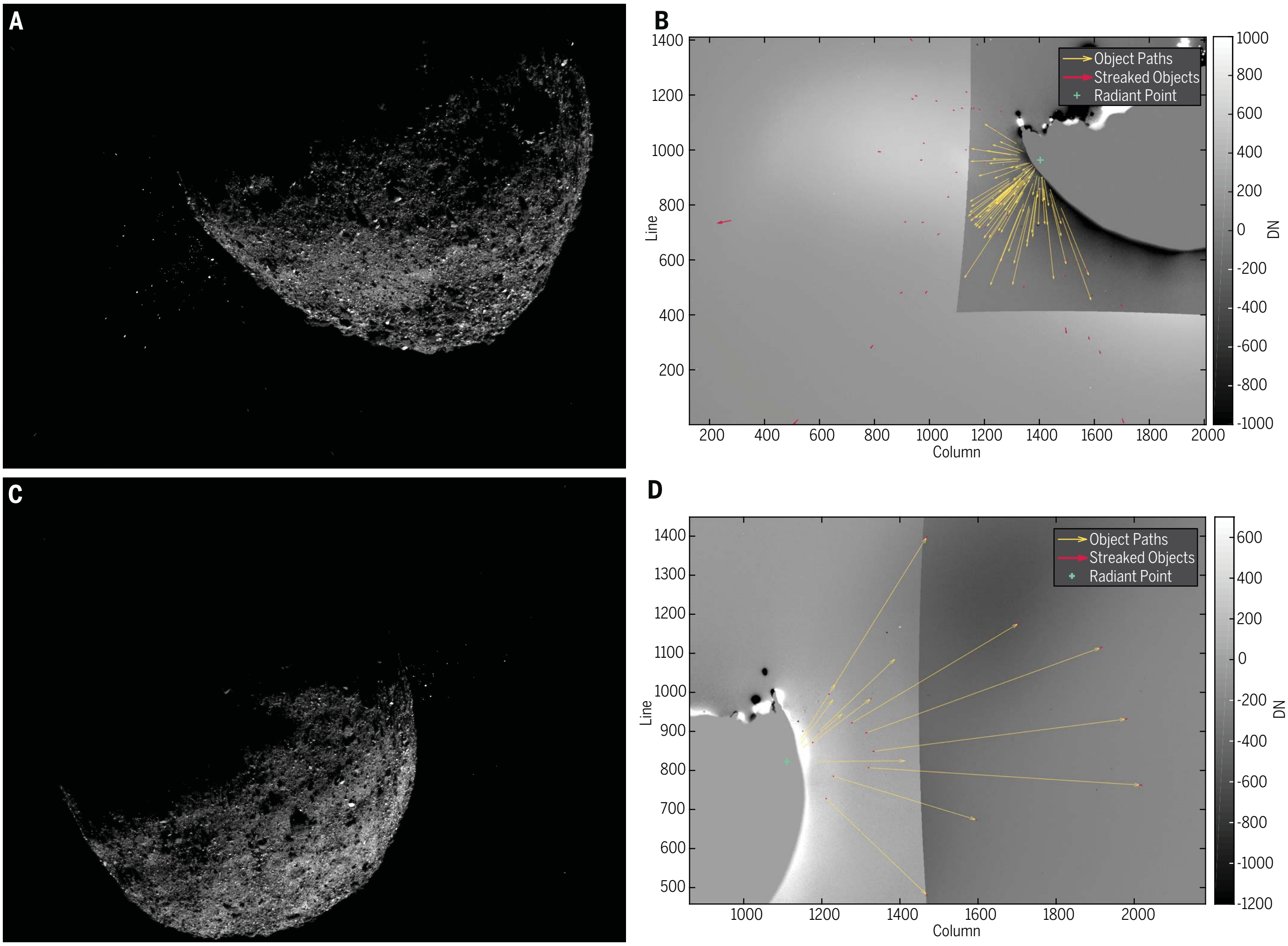


Fig. 1. Particle ejections from Bennu. (**A** and **C**) Composite views of particle ejections from the surface of asteroid Bennu on (A) 6 January and (C) 19 January. These images were produced by combining two exposures taken by the NavCam 1 imager in immediate succession: a short-exposure image (1.4 ms), showing the asteroid, and a long-exposure image (5 s), showing the particles (*12*). Image processing techniques were applied to increase the brightness and contrast of the ejected particles, which would otherwise be invisible at the same time as the bright asteroid surface (*12*). The original images are shown in fig. S3. In (A), Bennu's north (+*z*) pole is to the top right, pointing into the image; the subobserver latitude is −36°. In (C), the north pole is to the top right, pointing out of the image; the subobserver latitude is 60°. (**B** and **D**) Two NavCam images taken immediately after the ejection events on (B) 6 January and (D) 19 January are registered on the center of Bennu and differenced to highlight any moving particles. Particles moving at high velocity appear as streaks in a single image (red) that provide position information at the start and end of the exposure. The paths of particles moving more slowly (yellow) are identified from individual particles detected in the earlier image that also are present in the later image, farther from Bennu's limb. For each event, the apparent motion of the individual particles traces back to a radiant point on Bennu's surface (light blue cross) that indicates the potential source region on the near side of Bennu. A second possible source region occurs on the far side of Bennu, out of view. The shaded area closest to the asteroid [darker shading in (B), lighter shading in (D)] corresponds to where the two images share a common field of view and are differenced. The opposite shading corresponds only to the image with the larger field of view [earlier image in (B), later image in (D)]. DN, data number.

later shows objects in common with the earlier image that have moved away from Bennu, implying the movement of discrete particles (Fig. 1B). This observation triggered an immediate risk assessment of whether it was safe for the spacecraft to remain in orbit, which was concluded affirmatively, and led to an observational campaign to detect and characterize Bennu's apparent activity.

We increased the imaging cadence in response to the initial event to better characterize the frequency of particle ejections and any persistent particle population (table S1). Starting on 11 January, NavCam 1 began collecting image pairs of each field every 30 min. On 28 January, we again increased the cadence, collecting image pairs of each field every 20 min. This imaging frequency continued until 18 February. During this time period, we detected two additional ejection events of a similar scale, on 19 January (Fig. 1, C and D, and fig. S3) and 11 February (fig. S4). The distance from the spacecraft to Bennu's center of mass was 1.66 km for 6 January, 1.99 km for 19 January, and 1.64 km for 11 February. We used the imaging dataset to characterize these three events, which were the largest observed (they had the highest number of detected particles). We also observed several smaller events, in which fewer than 20 particles were detected (Fig. 2). There is also a persistent background level of particles in the Bennu environment; we detected a few particles per day during Orbital A, with observed increases immediately after the 19 January and 11 February events (Fig. 2).

Characterization of the largest observed events

For the 6 January, 19 January, and 11 February events, a particle distribution pattern near the limb of Bennu in the first image of each event is also apparent in the image collected ~7 min later, farther from the limb and dispersed (Fig. 1, B and D), and also appears in subsequent images for the 19 January and 11 February events. Using OpNav techniques developed for spacecraft navigation, we associated individual particle detections from this pattern and determined the trajectory and velocity of each particle (*12*). Fast-moving particles cross multiple pixels in a single exposure and appear as trails, providing position and velocity information within one image. For each event, OpNav analysis constrains two possible locations (a near and far radiant) on Bennu's surface from which the particles originated (Fig. 3, Table 1, and table S2) (*12*).

The 6 January event is the least constrained (particles detected in only two images) of the three largest events. We determined that the event originated at a high southern latitude (between about 57°S and 75°S) (Table 1 and Fig. 3A) (*12*), with an ejection time between 15:22 and 16:35 local solar time (LST). However, the event location relative to the spacecraft and the limited dataset make estimating the precise latitude and ejection time difficult. For this event, we determined speeds for 117 of the 200 observed particles, ranging from 0.07 to 3.3 m s^{-1}. Fifty-two particles were moving more slowly than Bennu's escape velocity [20 cm s^{-1} for the volume-averaged Bennu radius (*12*, *16*)] (fig. S5).

Because of the increased imaging cadence, there is a more extensive dataset for the 19 January and 11 February events. We used the output of the OpNav characterization to provide initial conditions for higher-fidelity orbit determination (OD) modeling. In these models, we assumed that the particles from a given event left Bennu's surface at the same location on a trajectory influenced by point-mass gravity (*12*). We performed this analysis on 24 particles from the 19 January event (Movie 1) and 25 particles from the 11 February event. For these two events, with individual particles identified in more than three images, this analysis allows us to estimate a single location for the particle source location (Fig. 3, B and C) as well as ejection timing and initial velocity vectors (Table 1).

We determined the ejection epoch (moment in time) by extrapolating the OD solutions backward to the point where they intersect Bennu's surface. This analysis shows that the event on 19 January occurred at 00:53:41 ± 4 s (3σ) UTC from a location on Bennu at latitude 20°N, longitude 335°. The epoch corresponds to 16:38 LST at that location. Surface ejection velocity magnitudes ranged from 0.06 to 1.3 m s^{-1}. The 19 January timing data show a bimodal distribution, with a small peak occurring 6 min before the main epoch (fig. S6), suggesting that some of the particles may have ejected in a smaller event before the large release. The event on 11 February occurred at 23:27:28 ± 6 s (3σ) UTC from latitude 20°N, longitude 60°, corresponding to 18:05 LST, with observed velocity magnitudes ranging from 0.07 to 0.21 m s^{-1}. All particles from this event appear to have left the surface nearly simultaneously (fig. S6). Many of the characterized particles are on ballistic trajectories that reimpact the surface on the night side of Bennu, whereas high-velocity particles escape on hyperbolic trajectories (Movie 1).

Images of the particle source locations on Bennu (Fig. 3, A to C) show no obvious geological distinction from other locations on the surface of Bennu. The event radiant locations contain abundant rocks that are diverse in size and surface texture, as well as small circular depressions that may be impact craters. However, similar features are globally distributed on Bennu (*17*, *18*). We analyzed the normal albedo distribution of the two better constrained source regions (19 January and 11 February) and found that they are similar to the global distribution for Bennu (*19*), averaging 0.042 ± 0.003 (1σ) (Fig. 3, D and E) (*12*). The lack of obvious morphologic or albedo variation may be due to the very low energies associated with the ejection events (Table 1 and table S3).

Characterization of gravitationally bound particles

In addition to particles released in ejection events, we observed a gravitationally bound background population of particles in the Bennu environment (Fig. 2). Among these are a few objects that remain in orbit for several days. From among the 215 tracks (linkages of individual detections of the same particle over a short time), we identified a representative group of six distinct gravitationally bound particles for further analysis. The trajectories around Bennu of these six particles and their altitude histories are shown in Fig. 4. Orbital elements are given in table S4 and fig. S7. Particles 1 to 4 are on short-lived orbits, persisting for 4 to 17 revolutions, with lifetimes ranging from 2 to 6 days. These orbits show a range of inclinations, from near equatorial to polar. Both prograde and retrograde orbits occur. The semimajor axis of particle 1 is >1 km, compared with 0.4 to 0.5 km for particles 2 to 4. Particles 5 and 6 are suborbital. By extrapolating the orbits back to the time when they intersected Bennu's surface, we determined that three of the six particles ejected from the night side of Bennu (between 18:00 and 06:00 LST) (table S5). The six particles were ejected with orbital velocities in the range of 15 to 20 cm s^{-1}. Surface-relative velocities at ejection range from roughly 10 to 25 cm s^{-1}.

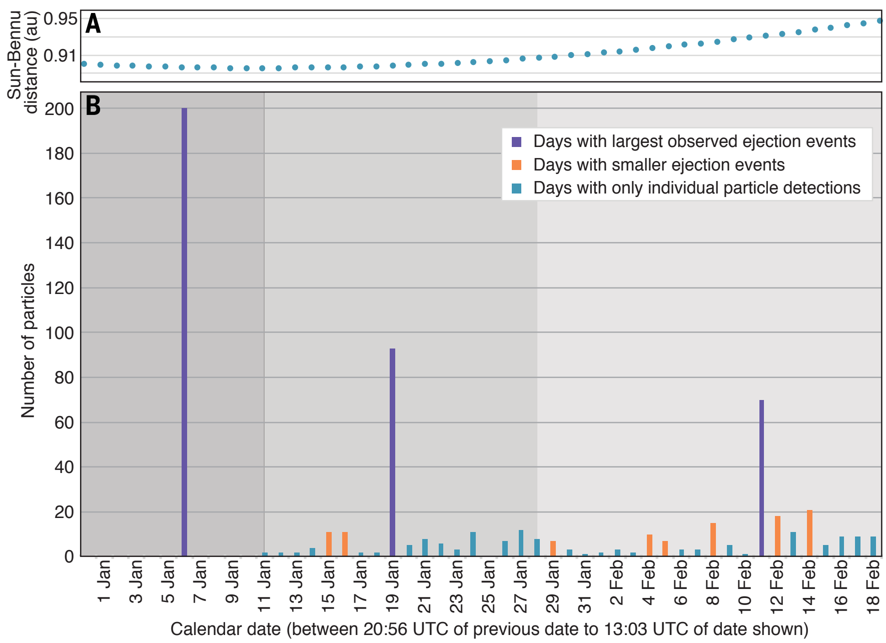
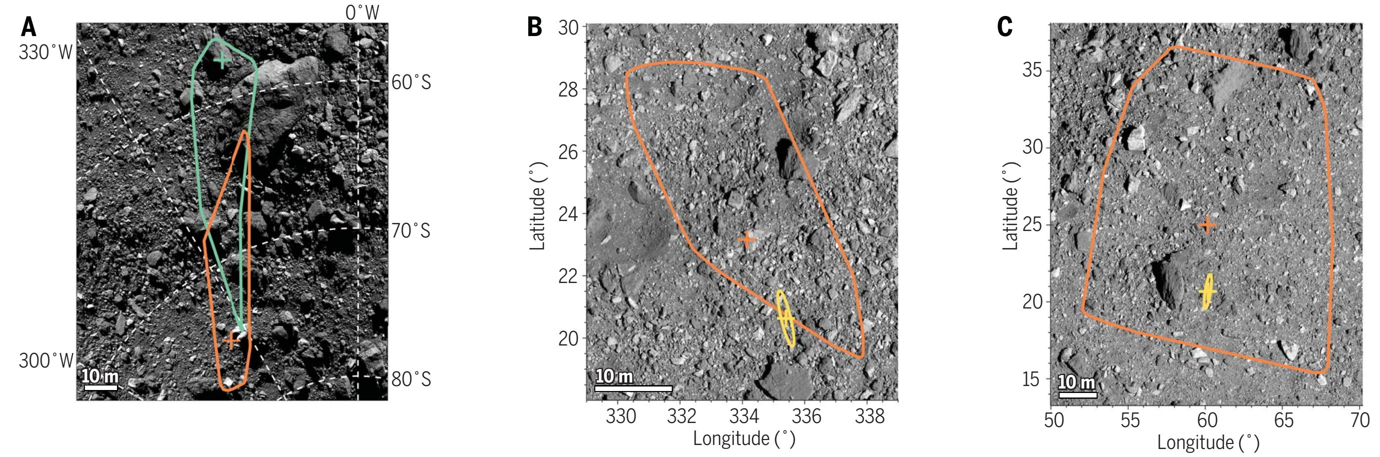


Fig. 2. Particle detections during the Orbital A mission phase. (**A**) Distance of Bennu from the Sun over the same time period as shown in (B). (**B**) Particle detections associated (purple and orange) and not associated (light blue) with observed ejection events. The changes in the background shading indicate when the observation cadence increased on 11 January and again on 28 January 2019.



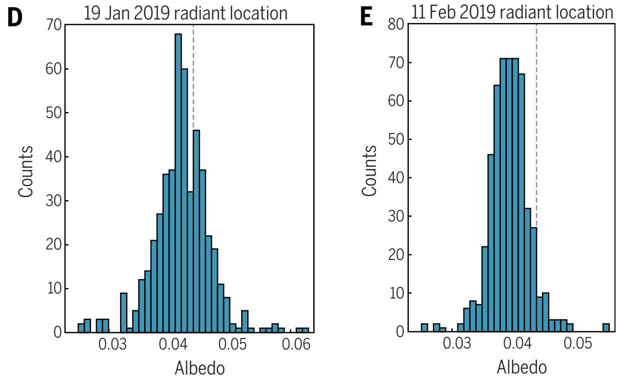


Fig. 3. Ejection source regions and their surface albedo distributions. (**A** to **C**) Radiant locations of the (A) 6 January, (B) 19 January, and (C) 11 February particle ejection events are overlain on a mosaic of Bennu (*12*). Orange and teal crosses indicate the far and near candidate radiant locations, respectively, determined from OpNav analysis; orange and teal outlines enclose the 3σ uncertainty region. The yellow crosses indicate the most likely source location determined from OD analysis; yellow lines trace the 3σ uncertainty. (**D** and **E**) For the locations of the latter two events, which are more tightly constrained, we show the surface albedo distributions (radiant locations with 3σ uncertainties) (*12*). The dashed vertical lines indicate the average albedo of Bennu's surface (*19*).

Particle properties

We constrained the area-to-mass ratios (where area is cross-sectional) of the six bound particles by using the trajectory information and modeling the nongravitational forces, which primarily arise from radiation pressures (table S5) (*12*). The particle trajectories also enabled us to calculate the phase angle and range of each observation to the spacecraft, from which we determined the photometric phase functions for particles 1 to 3, constraining the visible absolute magnitude of each particle (table S5). Combining the area-to-mass ratio and absolute magnitude information, and assuming a spherical shape, defines a distinct curve in density (ρ)–albedo (p_V) space for each particle (fig. S8). If we further assume particles with densities of 2 g cm^{-3} [on the basis of Bennu meteorite analogs (*20*)], then their normal albedos range from 0.05 to 0.3. In that case, the derived albedos are brighter than 96% of the material on Bennu, and the particle diameters range from 0.4 to 4.4 cm. If, on the other hand, the particles have normal albedos of 0.04, which is consistent with the average surface material on Bennu (*19*), then the densities range from 0.7 ± 0.3 (1σ) to 1.7 ± 0.4 (1σ) g cm^{-3} (fig. S8). The high end of this range is consistent with meteorite analogs. The lower densities lead to larger particle diameters, ranging from 1.2 to 8.5 cm. Given these uncertainties, we conclude that the particle diameters are in the range of <1 to ~10 cm.

With these constraints on the particle sizes, and the ejection velocities from the OD analysis, we can estimate the energy of the ejection events (Table 1 and table S3) (*12*). Such estimates should be considered lower limits because we may not have observed all ejected particles. In addition, our calculation assumes that the ejected particles had the average surface albedo of Bennu (0.044) (table S3) and the meteorite analog density of 2 g cm^{-3}. For 6 January, the 124 particles with measured photometry ranged in size from <1 to 8 cm, yielding a minimum event energy of ~270 mJ. For the 19 January event, more than 93 photometrically measured particles with radii between <1 and 7 cm ejected from the surface, giving a minimum event energy of 100 mJ. For 11 February, more than 60 particles with radii between <1 and 7 cm ejected from the surface, with an associated minimum event energy of 8 mJ (uncertainties on the event energies are provided in Table 1).

Possible ejection mechanisms

Several constraints apply to the particle ejection mechanism: The three largest observed ejection events occurred in the late afternoon, between 15:22 and 18:05 LST. The largest observed event (6 January) occurred days before Bennu reached perihelion (Fig. 1). The particles left the surface at discrete times. The observed particles ranged in size from <1 to ~10 cm. The ejection locations occurred over a range of latitudes from 75°S to 20°N. Particle velocities ranged from 0.07 to at least 3.3 m s^{-1}. The minimum kinetic energy of the ejected particles ranged from 8 to 270 mJ, assuming that the particles have albedos equivalent to the surface average of Bennu. Smaller events occurred that ejected fewer than 20 observed particles. Individual particles were ejected at a range of local solar times, including at night.

Dust ejection is a common phenomenon in comets and active asteroids. Even for well-studied comets such as 67P/Churyumov-Gerasimenko, substantial uncertainty exists as to the physical mechanism through which particles are released from the surface (*21*). We consider multiple hypotheses for the particle ejection mechanism, evaluating their respective strengths and weaknesses. These include rotational disruption, electrostatic lofting, comet-like ice sublimation, phyllosilicate dehydration, thermally driven stress fracturing, meteoroid impacts, and secondary impacts.

Rotational disruption

Mass shedding or splitting that results from rotational instability has been identified as a

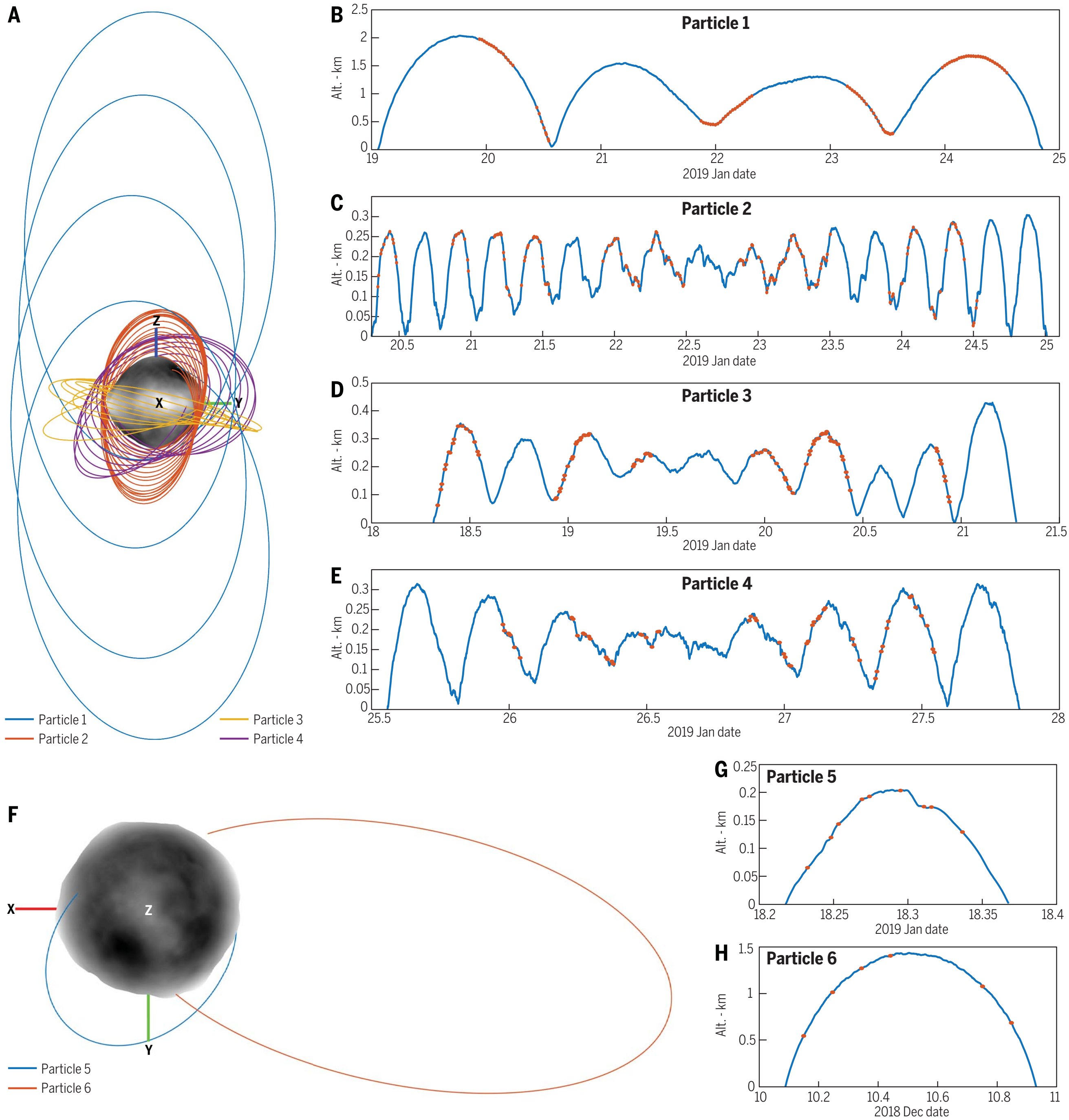


Fig. 4. Gravitationally bound particle altitudes and trajectories. (**A**) Trajectories and (**B** to **E**) altitude above Bennu's surface over time for four orbiting particles (particles 1 to 4) (fig. S7 and tables S4 and S5). On the altitude plots, red circles mark the times of observations used in the trajectory estimate. Axis scales of the altitude plots differ. For particles 2 and 3, it is not clear whether the last revolution depicted occurred or whether the particle impacted at the previous periapsis passage. The ragged appearance of the curves is a result of the rough topography of the surface of Bennu. (**F**) Trajectories and (**G** and **H**) altitude above Bennu's surface over time, as in (A) to (E) but for two suborbital particles (particles 5 and 6) (fig. S7 and tables S4 and S5). The trajectories are seen from above Bennu's north pole [*x* axis toward the Sun, *z* axis close to Bennu's north (positive) pole, *y* axis roughly in the direction of Bennu's heliocentric velocity]. Particle 6 is the earliest evidence of a particle in imaging data (10 December).

possible explanation for the activity of the smaller active asteroids (22). In this scenario, rapidly rotating asteroids experience centrifugal forces greater than the centripetal forces from self-gravity, leading to particle ejection preferentially from low latitudes. Particles launched from Bennu's surface would have a maximum velocity equal to the equatorial surface velocity of 10 cm s^{-1} (on the basis of Bennu's ~250-m equatorial radius and ~4.3-hour rotation period). This mechanism would preferentially produce particles in equatorial orbits in the rotational direction. It is not capable of launching particles on retrograde or hyperbolic trajectories, as we observed.

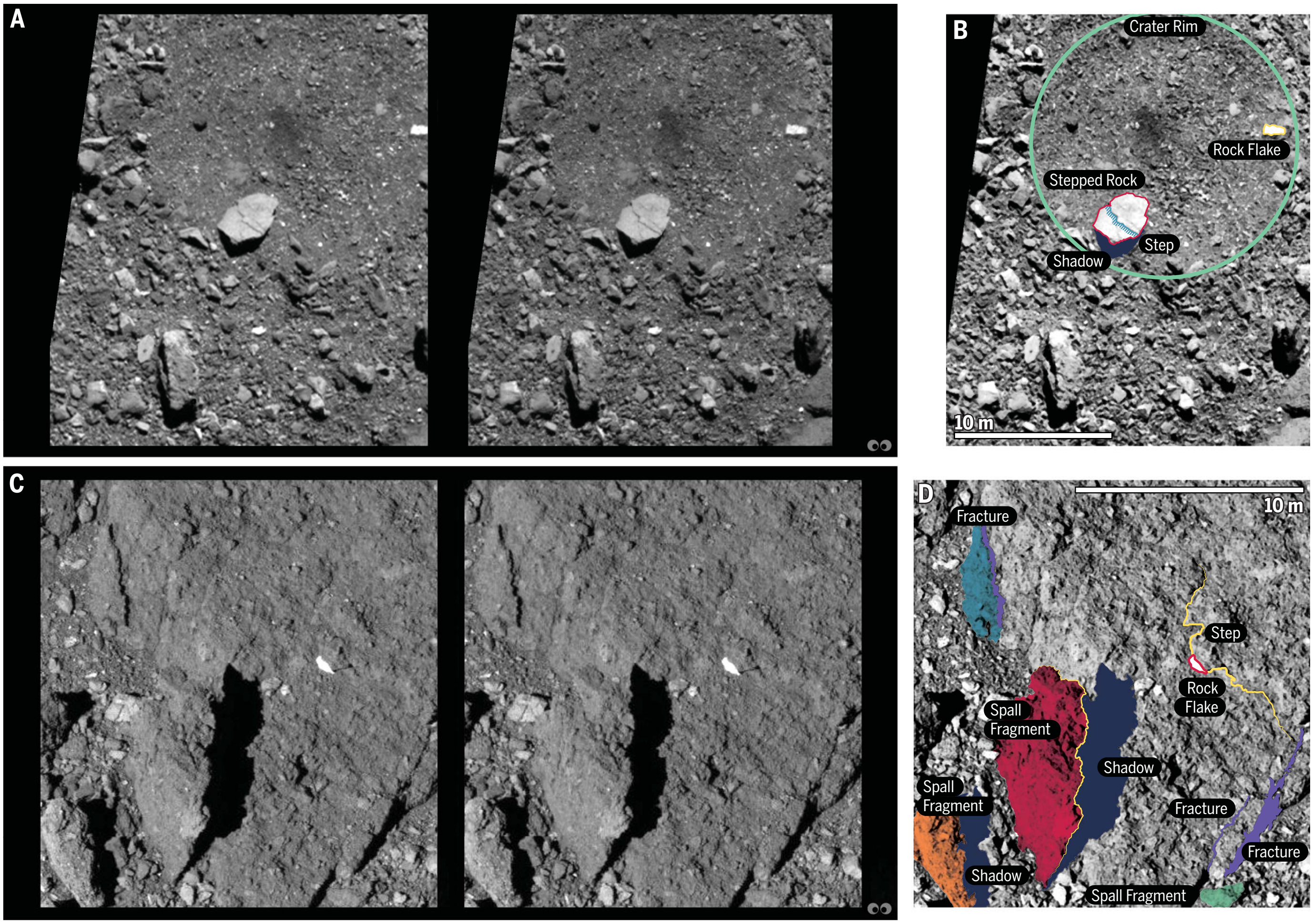


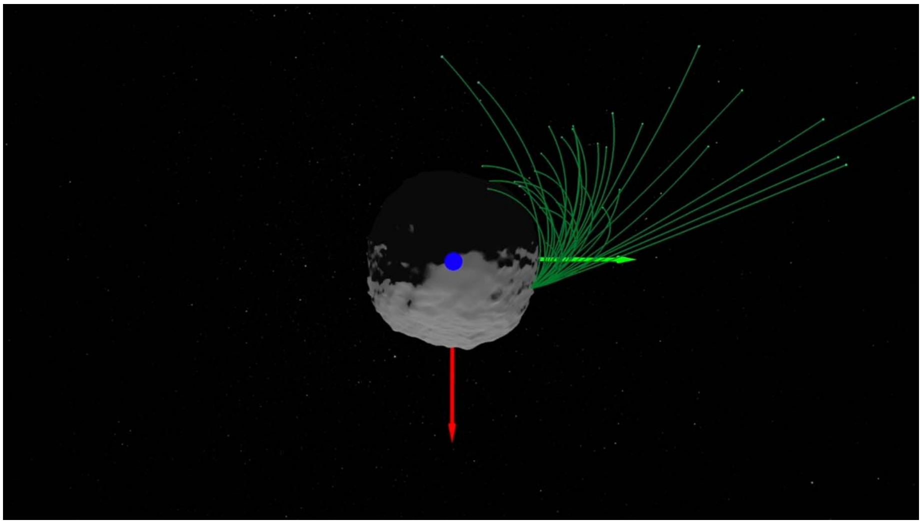
Fig. 5. Two distinct types of exfoliation textures on Bennu. In all images, north on Bennu is down. The PolyCam telescopic imager (*12*) acquired the four frames in (A) and (C) while the spacecraft moved with respect to the surface at a speed of 9 cm s^{-1} with exposures of (A) 1/300 of a second and (C) 1/200 of a second. These side-by-side stereo images are presented in the stereo "cross-eyed" configuration. A stereoscope-viewing version is available in fig. S10. Each pair of images has been adjusted to match their brightness, contrast, and shadow positions. (**A**) The parallax angle between these two images is 12°. Phase angle, 44°; pixel scale, 6.6 cm per pixel; (longitude, latitude), (90°, 11°). (**B**) Annotated version of the image on the right in (A). The large, 5-m white rock on the crater rim displays a flat face, with a well-defined step crossing its center. A white "flake" is present in the upper right. (**C**) The parallax angle between these two images is 8°. Phase angle, 30°; pixel scale, 4.7 cm per pixel; (longitude, latitude), (44°, −30°). (**D**) Annotated version of the image on the right in (C). The large black boulder displays exfoliation textures along both the east and west faces, with fractures running parallel to the texture in the rock. The large rock column in the bottom left has a profile that matches that of the step in the boulder, suggesting that this fragment may have been uplifted in an energetic exfoliation event. Even though the rock slab measures 5 by 5 by 1 m, it would only require ~5 J of energy to lift it, assuming a density of 2 g cm^{-3}. Other spalled fragments are present around the base of the large boulder.

Electrostatic lofting

Electrostatic lofting is the phenomenon of dust particles detaching from a surface once the electrostatic force on the particles exceeds those of gravity and cohesion (which bind the particles to the surface). The surface of an airless body (such as the Moon or an asteroid) interacts directly with the solar wind plasma, which charges the particles and produces a near-surface electric field. The electrostatic force is the product of the grain charge and the local electric field. Although electrostatic lofting has been discussed as a possible mechanism of the lunar horizon glow (*23*), when considering cohesion, there remained a discrepancy between the electrostatic force necessary to loft particles and the charging conditions hypothesized to be present in situ (*24*). Charge exchange between individual particles may produce very strong, short-scale electric fields that are capable of lofting particles in microgravity environments (*25*, *26*). It is possible to electrostatically loft particles up to millimeters in radius at small asteroids such as Bennu (*27*), smaller than those we observed. The velocities of electrostatically lofted particles are likely to be less than 1 m s^{-1}, unless additionally accelerated away from the surface by solar radiation pressure (*27*).

Ice sublimation

Dust release from comets is a major source of interplanetary dust particles. On comets, ice sublimation results in gas drag forces that eject dust particles from the surface (*21*). The gas-drag forces accelerate the released dust within a few times the radius of the nucleus, until solar radiation pressure takes over. For such sublimation to be the driver of the Bennu events, ice must be present at or near the surface. Several observed ejection events occurred at relatively low latitudes, where temperatures reach ~390 K (*17*). At these temperatures, major cometary ice species (CO, CO_2, and H_2O) are not stable [for example, (*28*)]. Additionally, there are no water-ice absorption features at 1.5 or 2.0 μm in spectra of the surface (*20*). Subsurface ice could be trapped at depths greater than 1 m at some locations for long periods (*29*). Rapid volatile release from such a reservoir would require exposure by large



Movie 1. Animation showing the results of the orbit analysis for a subset of the particles ejected from Bennu on 19 January. The highest-velocity particles are on escape trajectories and leave the Bennu environment. Most of the particles are on suborbital trajectories and reimpact the surface, primarily on the night side of the asteroid.

impacts or deep thermal cracking at meter scales. We observed no geologic evidence of such processes acting recently at the event locations (Fig. 3). There is also no evidence of a coma or jets associated with volatile release (Fig. 1 and figs. S3 and S4).

Phyllosilicate dehydration

Although ice has not been observed on Bennu, the surface is rich in water-bearing minerals. Spectroscopy has shown that the surface is dominated by hydrated phyllosilicates, with the closest spectral match being CM-type carbonaceous chondrite meteorites (*20*). Evolved gas analysis experiments on Murchison (a CM chondrite) have demonstrated that considerable volatile release can occur when heated from ambient temperature up to 473 K under vacuum [for example, (*30*–*32*)]. Although this temperature is ~70 K higher than the peak temperatures on Bennu, such low-temperature water release from Murchison indicates that the thermal dehydration of minerals begins with the loss of weakly bound adsorbed and interlayer water.

Mechanical stresses on Bennu's surface may generate adsorbed water, such as that released in laboratory experiments. The CM chondrites are dominated by Mg-rich serpentine and cronstedtite, an Fe-rich phyllosilicate [for example, (*33*)]. In these hydrated phases, particle size reduction through grinding enhances dehydroxylation and yields highly disordered material (*34*). The dehydroxylation reaction is substantially accelerated owing to the transformation of structural hydroxyls into adsorbed water in the resulting matrix. If mechanical stresses on Bennu result in a similar chemical transformation, the structural OH component of the phyllosilicates that dominate the surface mineralogy may be converted into absorbed water concentrated within an outer layer of the surface rocks. It is possible that the release of this adsorbed water within cracks and pores in boulders could provide a gas pressure leading to disruption of rock faces, such as is thought to occur on near-Earth asteroid (3200) Phaethon (*35*).

Meteoroid impacts

Solid bodies in space are routinely impacted by a steady flux of small meteoroids. Because Bennu is on an Earth-like orbit, we expect the flux of meteoroids at Bennu's surface to be similar to that on Earth, once corrected for gravitational focusing. A model of the interplanetary dust flux in near-Earth space has been determined by using data from in situ spacecraft measurements and lunar microcrater studies (*36*) and is widely adopted for meteoroid flux in near-Earth space (*37*). Lunar meteoroids typically impact at velocities between 13 and 18 km s^{-1} (*38*). If we assume an average velocity of 15.5 km s^{-1} for meteoroids at Bennu, an impact by an interplanetary dust particle with mass 2.5 μg would deposit 300 mJ of energy into the surface, which is consistent with the estimated energy of the largest observed event (6 January). However, Bennu has a cross-sectional area of 1.96×10^5 m^2; applying this value to the interplanetary dust flux model (*36*), we found that Bennu should be hit by a particle of this size on average once every minute, which is much more frequently than the observed ejection cadence. The large ejection events occurred on a roughly 2-week cadence. At that frequency, Bennu should be hit by an average of one meteoroid with a mass ~3000 μg, depositing more than 360,000 mJ of energy into the surface if it impacted at 15.5 km s^{-1}. Thus, only 0.07% of the impact energy from such events would need to be transferred to the particles to produce the largest observed ejection event.

The result of hypervelocity impacts into Bennu's surface depends substantially on the mass and velocity of the impacting grain and on the strength of the target material. Particle impacts at velocities on the order of 2.5 to 3 km s^{-1} produce well-developed craters with rims, fracturing, and spallation of a large number of particles (*39*). At higher speeds, such impact events produce little ejecta; instead, they deposit energy into a small volume of the asteroid surface, causing melting, vaporization, and at the highest energy densities, ionization of the target and impactor material producing plasma (*40*, *41*). It is possible that the observed ejection events are the result of low-velocity meteoroid impacts, which occur much less frequently. Alternatively, the particles may be accelerated by the small fraction of impact energy from more frequent, high-velocity impacts that did not result in plasma production.

Thermal stress fracturing

Bennu's surface experiences extreme temperature variations over its 4.3-hour rotation period. Laboratory studies (*42*) showed that the CM chondrite Murchison quickly developed cracks and spalled particles from diurnal temperature cycling under near-Earth asteroid surface conditions. At the mid-latitudes, where the 19 January and 11 February events occurred, the surface temperature plunges to 250 K in the predawn hours and reaches a peak of 400 K at ~13:00 LST (*12*). Because Bennu has a moderate thermal inertia of 350 J m^{-2} K^{-1} $s^{-½}$ (*17*), the maximum temperature at the thermal skin depth (penetration depth of daily thermal conduction) of ~2 cm occurs later in the afternoon, at ~16:00 LST. The amplitude of temperature variation falls by a factor of *e* at one thermal skin depth. Thus, for a region on Bennu whose maximum surface temperature is 400 K, the peak temperature at a depth of ~2 cm reaches 325 K, inducing a strong thermal gradient over this short distance that cycles every 4.3 hours.

Thermal cycling can drive the growth of cracks in rocks over a range of spatial scales within the thermal skin depth, controlled by the amplitude and frequency of the temperature cycle, mineral composition, constituent grain size, the overall rock shape, and its orientation relative to the Sun. At the bulk scale, stresses associated with temperature gradients and surface cooling are induced in different regions of a boulder at different times throughout the thermal cycle. Stresses that arise in the

shallow interior of large boulders tend to drive surface-parallel crack propagation (43). In the thermal fatigue regime, subcritical crack growth occurs slowly, propagating fractures incrementally over many cycles. Crack propagation velocity increases with crack length, until catastrophic disruption occurs, which may disaggregate material and eject particles from the surface.

In terrestrial settings, thermal fatigue combined with tectonic unloading is known to cause rock dome exfoliation and energetic particle ejection (44). In these studies, rocks show the greatest evidence for stress and microfracturing in the afternoon and evening. Although the tectonic unloading effects, which are not likely to be present on Bennu, are thought to add to the energy in these events, much less energy is needed to eject particles in a microgravity environment. Such energy may be stored as a result of structural deformation related to thermal strain, providing excess energy that leads to particle ejection (35).

Secondary surface impacts

A possible mechanism for the small ejection events is the reimpact of disaggregated material released by larger events. Analysis of particle trajectories in the largest events show that the particles have a substantial velocity component in the direction of asteroid rotation. Because the largest events occur in the afternoon, a large fraction of the particles on suborbital trajectories impact the night side of the asteroid (Movie 1). During impact, these particles may bounce off the surface or collide with other small particles on the surface, resulting in subsequent ejection of a small number of low-velocity particles.

Dynamical calculations show that ejecta moving at surface-relative velocities up to 30 cm s^{-1} (escape velocity of ~20 cm s^{-1} plus Bennu's surface rotational velocity of 10 cm s^{-1}) lofted from the surface of Bennu can reimpact the surface days later (Movie 1). Depending on the impact location, reimpacting particles may be relaunched into a suborbital trajectory by bouncing off a hard surface such as a boulder (45) or ricocheting off a fine-grained surface (46, 47). Numerical simulations show that impacts on a fine-grained surface may result in the ejection of smaller surface particles at launch speeds that exceed the escape speed of Bennu (fig. S9). However, we have not directly observed particles ejecting from Bennu that are as large as the impactors in these simulations; in the energy regime that we have observed, particles of that size would not have traveled far enough from the asteroid to be detectable in our images. Our assessment thus leaves three viable candidates for the primary ejection mechanism: phyllosilicate dehydration, meteoroid impacts, and thermal stress fracturing (discussed in Conclusions and broader implications).

Evidence from Bennu's geology

Particle ejection from rock surfaces is consistent with the widespread observation of exfoliation features on Bennu's surface (Fig. 5 and fig. S10). Exfoliation is the division of a rock mass into lenses, plates, or parallel "sheets" because of differential stresses (48). For some bright boulders on Bennu (Fig. 5A), lineation is present on the rock faces, and they exhibit sheets that parallel the direction of fracture propagation. The more abundant dark boulders on Bennu also exhibit exfoliation (Fig. 5C). In these rocks, the exfoliation fractures are linear, but the finer-grained texture appears as blocky segments in the fracture profile. Spalled fragments are seen resting on the surface and lying around the base of dark boulders.

The observed textures are characteristic of surface stresses that drive surface-perpendicular cracking, segment exfoliation sheets, and cause near-surface disaggregation. We do not observe similar spalled fragments in the immediate vicinity of the brighter boulders. However, we observe bright rocks perched on the surfaces of boulders, in orientations that exhibit no evident alignment with the underlying boulder's texture (for example, the bright object in the center-right of Fig. 5C). These bright rocks tend to have plate-like morphologies, similar to the exfoliation textures observed on the flat surfaces of the brighter rocks. Thus, exfoliation and fracturing may be operating on all boulders on Bennu, but the response of the bright rocks may be different—ejecting material over large distances, even on hyperbolic escaping trajectories, whereas the darker boulders decompose on site, creating a halo of spalled fragments.

Implications for Bennu's geophysics

The existence of low-energy particle ejection events on Bennu may result in reimpacting particles preferentially concentrated within the boundaries of Bennu's rotational Roche lobe (the region where material is energetically bound to the asteroid surface, between latitudes of ~±23°) (16). A random distribution of ejection events with a sizable fraction of particle velocities less than the escape speed will preferentially transport material toward the equator owing to the lower geopotential. Once within the Roche lobe, the particles are trapped inside unless given a large enough speed (a few centimeters per second) and will be redistributed within the lobe owing to the chaotic orbital environment whenever lofted

Table 1. Characteristics of the three largest observed particle ejection events. The more extensive imaging datasets acquired for the 19 January and 11 February 2019 events, relative to that for the 6 January 2019 event, allowed higher-fidelity OD determination of the event locations and times. More detail is given in (12) and tables S2 and S3.

	6 January		19 January	11 February
Number of particles with photometry	124 (of 200 total observed)		93 (of 93 total observed)	60 (of 72 total observed)
Velocity range (m s^{-1})	0.07 to 3.3		0.06 to 1.3	0.07 to 0.21
Particle diameter range (cm, ±1σ)	<1 to 8 ± 3		<1 to 7 ± 3	<1 to 7 ± 3
Minimum event energy (mJ, ±1σ)	270 (+150/−225)		100 (+50/−85)	8 (+4/−7)
Event location	Near radiant	Far radiant	OD radiant	OD radiant
Latitude (degrees, ±3σ)	−74.95 (+12.65/−2.79)	−57.30 (+1.49/−17.49)	20.63 ± 0.30	20.68 ± 0.37
Longitude (degrees, ±3σ)	325.32 (+18.91/−10.28)	343.67 (+3.80/−14.73)	335.40 ± 0.09	60.17 ± 0.08
Local solar time (±3σ)	15:22 (+01:06, −00:36)	16:35 (+00:06, −01:05)	16:38:01 ± 00:00:23	18:05:31 ± 00:00:22
UTC time (±3σ)	20:58:28 ± 00:00:47	20:58:28 ± 00:00:47	00:53:41 ± 00:00:04	23:27:28 ± 00:00:06

(16). Particles ejected with higher energies that achieve orbit will preferentially reimpact in the equatorial region within the lobe because of the larger asteroid radius there [(16), figure 5 therein]. After impact (which occurs at low speeds relative to the escape velocity), the particles will not have sufficient energy to escape the Roche lobe and again will be preferentially trapped, leading to a concentration of returning particles in these regions, as opposed to a globally uniform distribution.

Previous observations have indicated a steady increase in Bennu's rotation rate that will lead to doubling of that rate in ~1.5 million years; this acceleration is consistent with the Yarkovsky–O'Keefe–Radzievskii–Paddack (YORP) effect (10, 49). The angular momentum associated with particles ejected on escaping trajectories could also influence the rotation rate. It is possible to generate the measured rotational acceleration of Bennu by ejecting several particles of diameter ~10 cm once per day in the westward direction from the equator, assuming no concurrent water vapor loss. A random ejection of escaping particles from the surface of a spinning body would produce a spin deceleration (50).

We summed the net angular momentum change from particles launched normal to every facet on the asteroid surface and given a sufficient ejection speed for escape (12, 16). We found that such a flux would always cause Bennu to spin slower (fig. S11), counteracting the YORP effect (50). This implies that the strength of the YORP effect on Bennu due to solar photons could be greater than originally estimated (10, 49). If Bennu were to eject, for example, on the order of 20 10-cm particles per day at a speed of 18 cm s^{-1} (the speed at which the effect is the greatest) normal to random points on its surface, then on average, its rotational acceleration would be slowed by less than 1% of the measured rotational acceleration. Thus, when averaged over the entire surface, the net effect of particle ejection is negligible relative to the YORP effect.

The linear momentum transfer from the particle ejections is orders of magnitude lower than that of the transverse acceleration because of thermal emission from Bennu, the operative component for the Yarkovsky effect (51). This acceleration peaks at ~10^{-12} m s^{-2} during perihelion (51). Such an acceleration leads to a daily change in velocity ΔV of 10^{-7} m s^{-1}, which is more than 7000 times the ΔV caused by a single 10-cm particle with a density of 2 g cm^{-3} escaping at 1 m s^{-1}.

Conclusions and broader implications

The ejection events on Bennu inform our understanding of active asteroids. There are substantial differences between active asteroids as commonly defined—where major mass loss events occur through processes such as large impacts, volatile release, and rotational acceleration, leading to mass shedding—and relatively small mass loss events as we see on Bennu. It is likely that there is a continuum of event magnitudes and that we have been limited to observing only the largest phenomena.

Mass loss observed during perihelion from the B-type near-Earth asteroid (3200) Phaethon, the parent body of the Geminids meteor shower, apparently consists of smaller particles [1 μm (52)] than observed at Bennu (<1 to ~10 cm). However, particles in the centimeter size range were not observable during studies of Phaethon at perihelion, and sub-centimeter particles would have been difficult to detect in NavCam 1 images. Particles in the millimeter size range are observed as Geminids meteors (53). The mass loss from Bennu between 31 December and 18 February (including the three largest ejection events characterized above) was ~10^3 g, which is orders of magnitude less than Phaethon's near-perihelion mass loss (~10^4 to 10^5 kg per perihelion passage) (7). The mass loss rate (~10^{-4} g s^{-1}) on Bennu is also many orders of magnitude less than the rates observed at other active asteroids (~10 to 10^3 g s^{-1}) (1). Mass loss as seen at Bennu suggests that Phaethon's current mass loss rate may include larger particles and be greater than remote observations imply.

Having evaluated multiple hypotheses for the mechanism of particle ejection on Bennu, we found that thermal fracturing, volatile release by dehydration of phyllosilicate rocks, and meteoroid impacts are plausible explanations. Rotational disruption and electrostatic lofting cannot explain the observed particle sizes and ejection velocities. There is no evidence for ice on the surface of Bennu or for recent exposure of a subsurface ice reservoir at the multiple ejection sites. Bennu's boulder morphology and the event ejection times are consistent with exfoliation as a result of thermal fracturing, phyllosilicate dehydration, or an interplay between these two mechanisms. Because we expect meteoroid flux to be greatest in the leading hemisphere (late afternoon on Bennu because of its retrograde rotation), the ejection event times are also consistent with meteoroid impacts. It is possible that multiple mechanisms operate in combination. Reimpacting particles could play a role in the smaller ejections or contribute to the larger events.

The particles that escape from Bennu on parabolic or hyperbolic orbits will escape onto heliocentric orbits, which we expect to disperse over time into a meteoroid stream. On the basis of the measured ejection velocities, meteoroids released after 1500 CE would not have spread wide enough to bridge the current distance between the orbits of Bennu and Earth, 0.0029 AU, but will do so when that distance decreases later in the 21st century (54). However, if Bennu was active in the past, and the ejected particles survive for thousands of years, planetary perturbations would spread the stream wide enough to cause an annual meteor shower on present-day Earth around 23 September. The shower would radiate from a geocentric radiant at right ascension 5°, declination −34°, and speed 6.0 km s^{-1} (54), corresponding to an apparent entry speed of 12.7 km s^{-1} (12). Meteoroids moving this slowly would create meteors of integrated visual magnitude +2 to −5, assuming an 0.7% luminous efficiency (55). The stream would not easily blend with the sporadic background over thousands of years. No shower is detected in current meteor orbit survey data (56), but those data have poor coverage in the Southern Hemisphere.

The primary objective of OSIRIS-REx is to return samples of centimeter-scale rocks from the surface of Bennu to Earth for analysis (14). We have observed centimeter-scale particles frequently being ejected and reimpacting the asteroid surface. It is possible that the collected sample will contain some particles that were ejected and returned to Bennu's surface.

REFERENCES AND NOTES

1. D. Jewitt, H. Hsieh, J. Agarwal, "The active asteroids", in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. of Arizona Press, 2015), pp. 221–241.
2. E. W. Elst, G. Pizarro *et al.*, Comet P/1996 N2 (ELST-PIZARRO). IAU Circular 6456 (1996).
3. D. Jewitt *et al.*, Hubble Space Telescope investigation of main-belt comet 133P/Elst-Pizarro. *Astron. J.* **147**, 117 (2014). doi: 10.1088/0004-6256/147/5/117
4. D. Jewitt *et al.*, Episodically active asteroid 6478 Gault. *Astrophys. J.* **876**, L19 (2019). doi: 10.3847/2041-8213/ab1be8
5. F. Moreno, J. Licandro, A. Cabrera-Lavers, F. J. Pozuelos, Early evolution of disrupted asteroid P/2016 G1 (PANSTARRS). *Astrophys. J.* **826**, L22 (2016). doi: 10.3847/2041-8205/826/2/L22
6. J. Li, D. Jewitt, Recurrent perihelion activity in, (3200) Phaethon. *Astron. J.* **145**, 154 (2013). doi: 10.1088/0004-6256/145/6/154
7. M.-T. Hui, J. Li, Resurrection of (3200) Phaethon in 2016. *Astron. J.* **153**, 23 (2017).
8. D. Jewitt, The active asteroids. *Astron. J.* **143**, 66 (2012). doi: 10.1088/0004-6256/143/3/66
9. D. S. Lauretta *et al.*, The OSIRIS-REx target asteroid (101955) Bennu: Constraints on its physical, geological, and dynamical nature from astronomical observations. *Meteorit. Planet. Sci.* **50**, 834–849 (2015). doi: 10.1111/maps.12353
10. C. W. Hergenrother *et al.*, The operational environment and rotational acceleration of asteroid (101955) Bennu from OSIRIS-REx observations. *Nat. Commun.* **10**, 1291 (2019). doi: 10.1038/s41467-019-09213-x; pmid: 30890725
11. M. G. Daly *et al.*, The OSIRIS-REx Laser Altimeter (OLA) investigation and instrument. *Space Sci. Rev.* **212**, 899–924 (2017). doi: 10.1007/s11214-017-0375-3
12. Materials and methods are available as supplementary materials.
13. B. J. Bos *et al.*, Touch And Go Camera System (TAGCAMS) for the OSIRIS-REx asteroid sample return mission. *Space Sci. Rev.* **214**, 37 (2018). doi: 10.1007/s11214-017-0465-2
14. D. S. Lauretta *et al.*, OSIRIS-REx: Sample return from asteroid (101955) Bennu. *Space Sci. Rev.* **212**, 925–984 (2017). doi: 10.1007/s11214-017-0405-1
15. B. Williams *et al.*, OSIRIS-REx flight dynamics and navigation design. *Space Sci. Rev.* **214**, 69 (2018). doi: 10.1007/s11214-018-0501-x
16. D. J. Scheeres *et al.*, The dynamic geophysical environment of (101955) Bennu based on OSIRIS-REx measurements. *Nat. Astron.* **3**, 352–361 (2019). doi: 10.1038/s41550-019-0721-3

17. D. N. DellaGiustina *et al*., Properties of rubble-pile asteroid (101955) Bennu from OSIRIS-REx imaging and thermal analysis. *Nat. Astron.* **3**, 341–351 (2019). doi: 10.1038/s41550-019-0731-1
18. K. J. Walsh *et al*., Craters, boulders and regolith of (101955) Bennu indicative of an old and dynamic surface. *Nat. Geosci.* **12**, 399 (2019). doi: 10.1038/s41561-019-0360-4
19. D. S. Lauretta *et al*., The unexpected surface of asteroid (101955) Bennu. *Nature* **568**, 55–60 (2019). doi: 10.1038/s41586-019-1033-6; pmid: 30890786
20. V. E. Hamilton *et al*., Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nat. Astron.* **3**, 332–340 (2019). doi: 10.1038/s41550-019-0722-2; pmid: 31360777
21. J.-B. Vincent *et al*., Local manifestations of cometary activity. *Space Sci. Rev.* **215**, 30 (2019). doi: 10.1007/s11214-019-0596-8
22. M. Drahus *et al*., Fast rotation and trailing fragments of the active asteroid P/2012 F5 (Gibbs). *Astrophys. J.* **802**, L8 (2015). doi: 10.1088/2041-8205/802/1/L8
23. J. J. Rennilson, D. R. Criswell, Surveyor observations of lunar horizon-glow. *Moon* **10**, 121–142 (1974). doi: 10.1007/BF00655715
24. C. M. Hartzell, D. J. Scheeres, The role of cohesive forces in particle launching on the moon and asteroids. *Planet. Space Sci.* **59**, 1758–1768 (2014). doi: 10.1016/j.pss.2011.04.017
25. X. Wang, J. Schwan, H.-W. Hsu, E. Grün, M. Horányi, Dust charging and transport on airless planetary bodies. *Geophys. Res. Lett.* **43**, 6103–6110 (2016). doi: 10.1002/2016GL069491
26. M. I. Zimmerman *et al*., Grain-scale supercharging and breakdown on airless regolith. *J. Geophys. Res. Planets* **121**, 2150–2165 (2016). doi: 10.1002/2016JE005049
27. C. M. Hartzell, Dynamics of 2D electrostatic dust levitation at asteroids. *Icarus* **333**, 234–242 (2019). doi: 10.1016/j.icarus.2019.05.013
28. E. L. Andreas, New estimates for the sublimation rate of ice on the Moon. *Icarus* **186**, 24–30 (2007). doi: 10.1016/j.icarus.2006.08.024
29. N. Schorghofer, Predictions of depth-to-ice on asteroids based on an asynchronous model of temperature, impact stirring, and ice loss. *Icarus* **276**, 88–95 (2016). doi: 10.1016/j.icarus.2016.04.037
30. E. K. Gibson Jr., Inorganic gas release studies and thermal analysis investigations on carbonaceous chondrites. *Meteoritics* **9**, 343–344 (1974).
31. I. L. ten Kate *et al*., VAPoR—Volatile Analysis by Pyrolysis of Regolith—an instrument for in situ detection of water, noble gases and organics on the Moon. *Planet. Space Sci.* **58**, 1007–1017 (2010). doi: 10.1016/j.pss.2010.03.006
32. E. K. Gibson Jr., S. M. Johnson, Thermogravimetric-quadrupole mass-spectrometric analysis of geochemical samples. *Thermochim. Acta* **4**, 49–56 (1972). doi: 10.1016/0040-6031(72)87062-X
33. D. S. Lauretta, X. Hua, P. R. Buseck, Mineralogy of fine-grained rims in the ALH 81002 CM chondrite. *Geochim. Cosmochim. Acta* **64**, 3263–3273 (2000). doi: 10.1016/S0016-7037(00)00425-7
34. A. Drief, F. Nieto, The effect of dry grinding on antigorite from Mulhacen, Spain. *Clays Clay Miner.* **47**, 417–424 (1999). doi: 10.1346/CCMN.1999.0470404
35. D. Jewitt, J. Li, Activity in geminid parent (3200) Phaethon. *Astron. J.* **140**, 1519–1527 (2010). doi: 10.1088/0004-6256/140/5/1519
36. E. Grün, H. A. Zook, H. Fechtig, R. H. Giese, Collisional balance of the meteoritic complex. *Icarus* **62**, 244–272 (1985). doi: 10.1016/0019-1035(85)90121-6
37. Committee for the Assessment of NASA's Orbital Debris Programs, Aeronautics and Space Engineering Board, Division on Engineering and Physical Sciences, *Limiting Future Collision Risk to Spacecraft: An Assessment of Nasa's Meteoroid and Orbital Debris Programs* (National Research Council, 2011).
38. H. A. Zook, "The state of meteoritic material on the Moon," in *Lunar Science VI*, 1301 (1975).
39. K. Fiege *et al*., Space weathering induced via microparticle impacts: 2. Dust impact simulation and meteorite target analysis. *J. Geophys. Res. Planets* **124**, 1084–1099 (2019). doi: 10.1029/2018JE005564
40. M. S. Thompson, R. Christoffersen, T. J. Zega, L. P. Keller, Microchemical and structural evidence for space weathering in soils from asteroid Itokawa. *Earth Planets Space* **66**, 89 (2014). doi: 10.1186/1880-5981-66-89
41. S. Drapatz, K. W. Michel, Theory of shock-wave ionization upon high-velocity impact of micrometeorites. *Z. Naturforsch. A* **29**, 870–879 (1974). doi: 10.1515/zna-1974-0606
42. M. Delbo *et al*., Thermal fatigue as the origin of regolith on small asteroids. *Nature* **508**, 233–236 (2014). doi: 10.1038/nature13153; pmid: 24695219
43. J. L. Molaro, S. Byrne, J.-L. Le, Thermally induced stresses in boulders on airless body surfaces, and implications for rock breakdown. *Icarus* **294**, 247–261 (2017). doi: 10.1016/j.icarus.2017.03.008
44. B. D. Collins *et al*., Thermal influences on spontaneous rock dome exfoliation. *Nat. Commun.* **9**, 762 (2018). doi: 10.1038/s41467-017-02728-1; pmid: 29472534
45. H. Yano *et al*., Touchdown of the Hayabusa spacecraft at the Muses Sea on Itokawa. *Science* **312**, 1350–1353 (2006). doi: 10.1126/science.1126164; pmid: 16741113
46. C. Maurel, P. Michel, J. Biele, R.-L. Ballouz, F. Thuillet, Numerical simulations of the contact between the lander MASCOT and a regolith-covered surface. *Adv. Space Res.* **62**, 2099–2124 (2018). doi: 10.1016/j.asr.2017.05.029
47. F. Thuillet *et al*., Numerical modeling of lander interaction with a low-gravity asteroid regolith surface: Application to MASCOT onboard Hayabusa2. *Astron. Astrophys.* **615**, A41 (2018). doi: 10.1051/0004-6361/201832779
48. G. R. Holzhausen, Origin of sheet structure, 1. Morphology and boundary conditions. *Eng. Geol.* **27**, 225–278 (1989). doi: 10.1016/0013-7952(89)90035-5
49. M. C. Nolan *et al*., Detection of rotational acceleration of Bennu using HST lightcurve observations. *Geophys. Res. Lett.* **46**, 1956–1962 (2019). doi: 10.1029/2018GL080658
50. A. R. Dobrovolskis, J. A. Burns, Angular momentum drain: A mechanism for despinning asteroids. *Icarus* **57**, 464–476 (1984). doi: 10.1016/0019-1035(84)90130-1
51. S. R. Chesley *et al*., Orbit and bulk density of the OSIRIS-REx target Asteroid (101955) Bennu. *Icarus* **235**, 5–22 (2014). doi: 10.1016/j.icarus.2014.02.020
52. D. Jewitt, J. Li, J. Agarwal, The dust tail of asteroid (3200) Phaethon. *Astrophys. J.* **771**, L36 (2013). doi: 10.1088/2041-8205/771/2/L36
53. V. Vojáček, J. Borovička, P. Koten, P. Spurný, R. Štork, Properties of small meteoroids studied by meteor video observations. *Astron. Astrophys.* **621**, A68 (2019). doi: 10.1051/0004-6361/201833289
54. Q. Ye, Prediction of meteor activities from (101955) Bennu. *Res. Notes AAS* **3**, 56 (2019). doi: 10.3847/2515-5172/ab12e7
55. R. J. Weryk, P. G. Brown, Simultaneous radar and video meteors—II: Photometry and ionisation. *Planet. Space Sci.* **81**, 32–47 (2013). doi: 10.1016/j.pss.2013.03.012
56. P. Jenniskens *et al*., CAMS: Cameras for Allsky Meteor Surveillance to establish minor meteor showers. *Icarus* **216**, 40–61 (2011). doi: 10.1016/j.icarus.2011.08.012

ACKNOWLEDGMENTS

We are grateful to the entire OSIRIS-REx Team for making the encounter with Bennu possible. **Funding:** This material is based upon work supported by NASA under Contracts NNM10AA11C and NNG13FC02C issued through the New Frontiers Program. A portion of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. OLA and funding for the Canadian authors was provided by the Canadian Space Agency. C.M.H., J.L.M., and P.T. acknowledge support from NASA's OSIRIS-REx Participating Scientist Program (grants 80NSSC18K0227, 80NSSC18K0239, and 80NSSC18K0280, respectively). P.M., G.L., and F.T. acknowledge funding support from the French Agency CNES and from the Academies of Excellence on Complex Systems and Space, Environment, Risk and Resilience of the Initiative d'EXcellence "Joint, Excellent, and Dynamic Initiative" (IDEX JEDI) of the Université Côte d'Azur. J.L. and J.d.L. acknowledge funding support from the projects AYA2015-67772-R (MINECO, Spain) and ProID20170112 (ACIISI/Gobierno de Canarias/EU/FEDER). B.Ro. acknowledges funding support from the Royal Astronomical Society (RAS) and the UK Science and Technology Facilities Council (STFC). **Author contributions:** D.S.L. led the scientific investigation and developed the hypotheses for ejection mechanisms. C.W.H. detected the particle ejection events and led the photometric modeling. C.K.M., J.N.K.Jr., and J.-Y.L. supported the photometric modeling efforts. S.R.C. led the team that performed the orbital element analysis of the six short-lived orbiting particles, supported by R.A.J., M.B., A.B.D., D.F., Y.T., W.M.O.Jr., D.J.S., and J.W.M. D.J.S. also calculated the Bennu Roche lobe and contribution of particle ejections to Bennu's observed rotational acceleration and Yarkovsky. M.C.M. led the team that performed the OpNav and OD analyses. J.M.L. performed the OD analysis for the three largest observed ejection events. J.Y.P. led the OpNav characterization of the three largest events, supported by A.J.L., E.J.L.-C., C.D.A., D.S.N., L.K.M., and E.M.S. M.G.D. is the lead instrument scientist for OLA and performed the analysis of the off-body lidar returns, supported by M.A.A. and J.A.S. B.J.B. is the lead instrument scientist for NavCam 1 and performed image calibration and image processing for the navigation images. B.Ri. and C.Y.d.'A. are the lead and deputy lead instrument scientists, respectively, for the OSIRIS-REx Camera Suite; they analyzed images of Bennu's surface for evidence of particle infall and processed the images used for the stereo pairs. D.N.D. is the lead image processing scientist for OSIRIS-REx and prepared the global mosaics used to register the particle source locations, supported by K.J.B., C.A.B., and D.R.G. K.J.B. also developed the NavCam camera model for use in registering NavCam images relative to the Bennu shape model and the background star fields. J.P.E. is the lead thermal analysis scientist for OSIRIS-REx, and B.Ro. developed the asteroid thermal model to determine the surface temperatures, skin depths, and thermal gradients at the particle ejection sites and globally across the asteroid. R.-L.B. developed the secondary impact hypothesis with support from P.M. and F.T. W.F.B. provided input on the meteoroid impact hypothesis and evaluated the other hypotheses in the context of the dynamical evolution of Bennu. H.C., J.d.L., and J.L. provided expertise on other known active asteroids. H.C.C.Jr. provided input on the potential mechanisms for ejection events and on the content of the manuscript. J.P.D. provided rock count data for testing the re-impacting particle hypothesis. D.P.G. counted rocks and provided input on EGA studies of meteorites and their low-temperature volatile release. C.M.H. developed the electrostatic lofting hypothesis. P.J. developed the hypothesis on the potential Bennu meteor shower. E.R.J. performed the geologic analysis of the particle ejection source regions. G.L. provided input on the relevance of thermal cycling experiments and regolith evolution. C.M. and B.M. identified and processed the stereo pair images. J.L.M. provided input on the feasibility of thermal fracture processes as a mechanism for particle ejection. H.L.R. led the development of the figures. S.H.S. provided software for processing NavCam images and identifying candidate particles. P.T. performed calculations for the statistical assessment of particles lifetime and fallback distribution. D.V. supported the development of the force model for nongravitational forces for orbiting particles, as well as giving input on the general implications and context of Bennu's activity. C.W.V.W. contributed substantively to the writing and preparation of the manuscript. **Competing interests:** H.C.C.Jr. is also affiliated with the Department of Earth and Planetary Science, American Museum of Natural History, New York, NY, USA. J.L.M. is also affiliated as a contractor with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. **Data and materials availability:** NavCam 1 images from Orbital A and OLA data from Preliminary Survey are available from the Planetary Data System (PDS) at https://sbn.psi.edu/pds/resource/orex/tagcams.html and https://sbn.psi.edu/pds/resource/orex/ola.html, respectively. The NavCam 1 images that we used are listed in data file S1. The parameters of the three largest ejection events are given in tables S2 and S3, and the derived orbital data for the six gravitationally bound particles are in tables S4 and S5 and data file S1.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/eaay3544/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S5
References (*57*–*92*)
Data File S1

11 June 2019; accepted 22 October 2019
10.1126/science.aay3544

RESEARCH ARTICLE

STEM CELLS

Stem cell–driven lymphatic remodeling coordinates tissue regeneration

Shiri Gur-Cohen[1], Hanseul Yang[1], Sanjeethan C. Baksh[1], Yuxuan Miao[1], John Levorse[1], Raghu P. Kataru[2], Xiaolei Liu[3], June de la Cruz-Racelis[1], Babak J. Mehrara[2], Elaine Fuchs[1]*

Tissues rely on stem cells (SCs) for homeostasis and wound repair. SCs reside in specialized microenvironments (niches) whose complexities and roles in orchestrating tissue growth are still unfolding. Here, we identify lymphatic capillaries as critical SC-niche components. In skin, lymphatics form intimate networks around hair follicle (HF) SCs. When HFs regenerate, lymphatic–SC connections become dynamic. Using a mouse model, we unravel a secretome switch in SCs that controls lymphatic behavior. Resting SCs express angiopoietin-like protein 7 (*Angptl7*), promoting lymphatic drainage. Activated SCs switch to *Angptl4*, triggering transient lymphatic dissociation and reduced drainage. When lymphatics are perturbed or the secretome switch is disrupted, HFs cycle precociously and tissue regeneration becomes asynchronous. In unearthing lymphatic capillaries as a critical SC-niche element, we have learned how SCs coordinate their activity across a tissue.

To replenish and repair the tissues in which they reside, stem cells (SCs) must not only self-renew but also generate differentiated lineages on demand (*1*). Their interactions with their microenvironment influence this decision (*2*–*4*). SC–niche interactions must be tightly regulated to avoid either excessive SC activity, which can cause tissue overgrowth and SC exhaustion, or insufficient activity, which can contribute to aging and defective tissue regeneration (*5*).

Despite the importance of niche constituents, SC–niche cross-talk is poorly understood. Additionally, most tissues have multiple niches, and the field still lacks an understanding of how SC niches are coordinated across a tissue. To tackle these problems, we used the murine hair coat as our paradigm because hair follicles (HFs) proceed through synchronized cycles of active growing (anagen) and resting (telogen) phases (the "hair cycle"), fueled by SCs within an anatomical niche ("bulge") (*3*, *6*) located just below the sebaceous glands (SGs) of every HF.

Hair growth is sensitive to systemic changes (*7*, *8*), hinting that vasculature (*9*) may be a key hair follicle stem cell (HFSC)–niche component. Although blood vessels are niche constituents in some tissues, whether and how lymphatic vasculature affects SC function is unclear (*10*–*16*). In this study, using cell biology, three-dimensional (3D) deep imaging, and molecular genetic approaches, we identify lymphatic capillaries as an intimate feature of the HFSC niche. We show that dynamic lymphatic remodeling, driven by SCs, regulates the regenerative process and integrates SC niches across the tissue.

[1]Howard Hughes Medical Institute, Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, NY 10065, USA.
[2]Department of Surgery, Plastic and Reconstructive Surgery Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [3]Center for Vascular and Developmental Biology, Feinberg Cardiovascular and Renal Research Institute, Northwestern University, Chicago, IL 60611, USA.
*Corresponding author. Email: fuchslb@rockefeller.edu

Lymphatic capillaries: A newly identified SC-niche component

Assessing vascular–SC spatial relationships was made possible by a recent clearing method that renders opaque tissue transparent while preserving cellular and subcellular tissue structures (*17*) (fig. S1A). Whole-mount immunofluorescence and 3D image reconstruction of skin exposed an array of dermal vessels, positive for panendothelial marker CD31, just below HF SGs. During telogen, large-diameter vessels closely approached keratin 24 ($KRT24^+$) HFSCs within the lower bulge (fig. S1B and movie S1).

HFSC-associated vessels were not blood vessels ($Endomucin^+$), but rather they were positive for both surface lymphatic vessel endothelial hyaluronan receptor-1 (LYVE1) and vascular endothelial growth factor tyrosine kinase receptor-3 (VEGFR3), establishing their lymphatic endothelial identity (Fig. 1, A and B; fig. S1, C to G; and movies S2 to S5). A similar association between lymphatic capillaries and the bulge was seen in human HFs, which spend most of their time in anagen (fig. S1H). We focused on mice, whose hair growth cycles are shorter and temporally choreographed. By studying telogen (Tel), it was clear that lymphatics were tightly associated with the bulge and, to a lesser extent, with progenitors (hair germ, HG) that are primed to undergo proliferation and fate commitment at the onset of tissue regeneration [anagen I (AnaI)] (*1*) (Fig. 1A and fig. S1I).

Lymphatic capillaries drain into collecting vessels, which differ molecularly and anatomically, as well as in terms of permissiveness to fluid and cell entry (*18*). Lymphatics associated with the HFSC niche were thin-walled and blind-ended lymphatic capillaries ($VEGFR3^+LYVE1^+$), whereas collecting vessels ($VEGFR3^+LYVE1^{neg}$) resided deeper within the dermis (Fig. 1, C and D, and fig. S1J). Moreover, relative to the posterior arrector pili muscle, capillaries were asymmetrically positioned anteriorly along each bulge (fig. S1K and movie S6), at sites where new SCs form during early anagen (*3*). During embryogenesis, connections between lymphatic capillaries and HFs also coincided with emergence of the quiescent bulge niche (*16*, *19*) (fig. S2 and movies S7 to S11).

Lymphatic capillaries maintain SC quiescence

To determine whether lymphatic integrity functions in HF quiescence, we perturbed the lymphatic vascular network. Using mice expressing *CreER* knocked into the *Prox-1* locus (*20*) and harboring *Rosa26-fl-stop-fl-YFP*, we first activated lineage-tracing in telogen skin and confirmed by immunofluorescence and flow cytometry that *Prox1-YFP*$^+$ cells were exclusively lymphatics (fig. S3, A to D). We then intercrossed a Cre-recombinase–inducible diphtheria toxin (DT) receptor line (*iDTR*) (*21*) with *Prox1-CreER* mice and induced iDTR expression during the extended second telogen.

A single DT intradermal injection induced lymphatic cell death and disrupted the network (Fig. 2A and fig. S3, E to H). This perturbation stimulated HFs to proliferate and enter anagen, irrespective of targeting lymphatics in first or second telogen or even early anagen (Fig. 2, B and C, and fig. S4). Anagen entry did not induce *Prox1* expression in HFSCs, blood capillaries remained intact, and the skin immune cell milieu resembled normal telogen-to-anagen transition (*22*) (figs. S4 and S5).

To further assess whether precocious HFSC proliferation might arise directly from disrupting the lymphatic–SC niche network, we administered soluble VEGFR3 receptor intradermally during second telogen. Because VEGFR3 signaling is essential for lymphatic endothelial cell proliferation and survival, its interception causes lymphatic regression (*23*). Notably, precocious HF anagen entry was recapitulated with this model (Fig. 2D and fig. S6). Given that VEGFR3 (*Flt4*) was expressed by lymphatics and not HFSCs (fig. S4D), these data further underscored a role for lymphatic capillaries in HF regeneration.

The lymphatic–SC niche is dynamic during physiological regeneration

Shortly after hair cycle onset (AnaII to AnaIII) in normal skin, $LYVE1^+PROX1^+VEGFR3^+$ lymphatic capillaries exhibited signs of remodeling

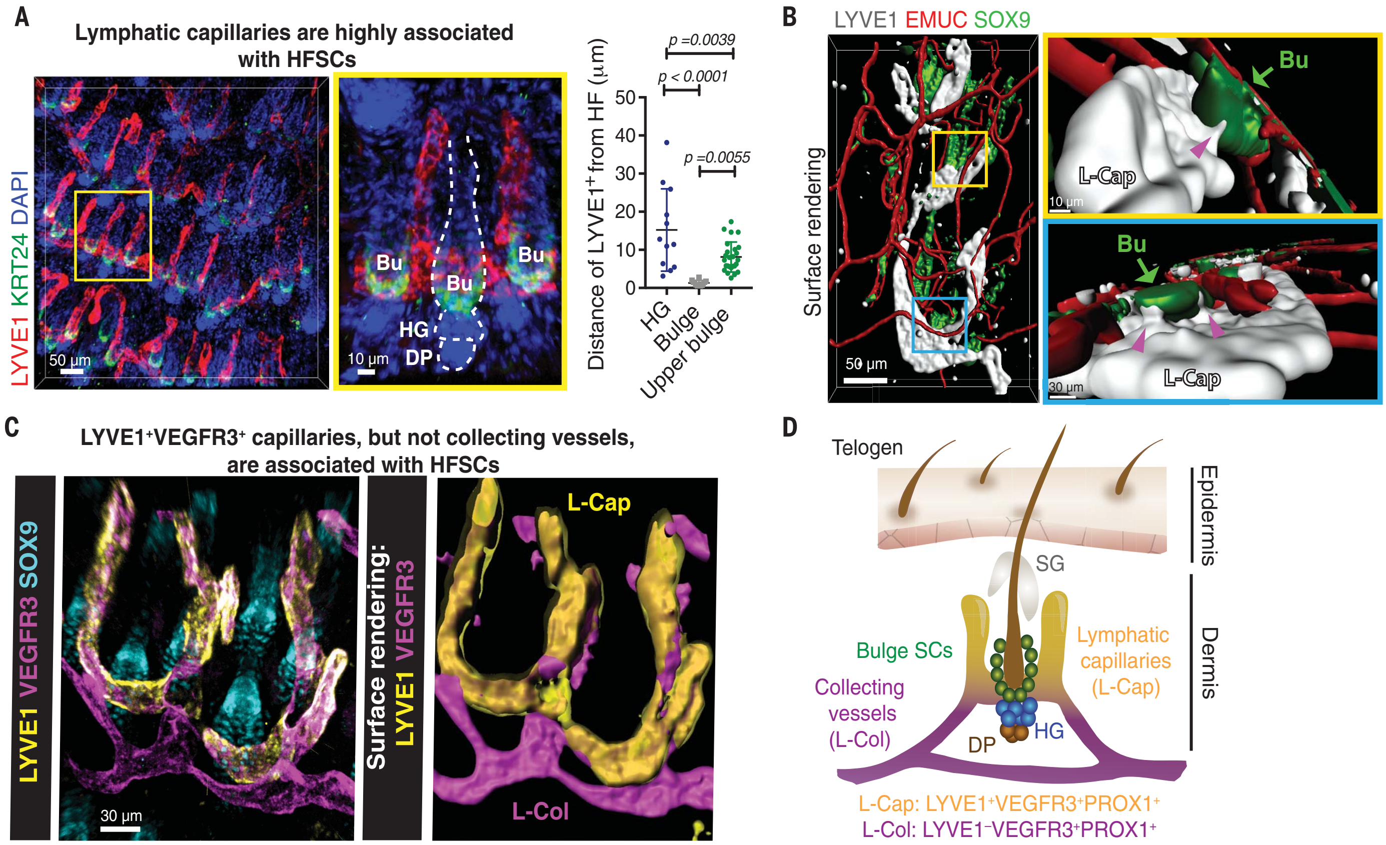


Fig. 1. Lymphatic capillaries tightly associate with HFSCs. (**A**) 3D image and quantifications demonstrating LYVE1^{+} lymphatic capillaries around telogen HF bulges (KRT24^{+}; boxed image enlarged at right) [n = 4 mice; multiple measurements per mouse; one-way analysis of variance (ANOVA); Tukey's multiple comparisons test]. DAPI, 4′,6-diamidino-2-phenylindole. (**B**) Surface renderings of SOX9^{+} bulges, LYVE^{+} lymphatic capillaries, and Endomucin^{+} (EMUC) blood capillaries (90°-angle views of boxed images are enlarged at right). (**C**) Lymphatic capillaries (L-Cap; LYVE1^{+}VEGFR3^{+}) but not collecting vessels (L-Col; LYVE1^{neg}VEGFR3^{+}) associate with telogen SOX9^{+} bulge SCs. (**D**) Schematic of SC–lymphatic association at telogen. Bu, bulge; HG, hair germ; DP, dermal papilla; SG, sebaceous gland.

around the SC niche (Fig. 3A and fig. S7A). Although capillaries still connected to underlying PROX1^{+}VEGFR3^{+}LYVE1^{neg} collecting vessels and maintained cell numbers with little or no signs of apoptosis, they were dissociated from SC niches and dilated (Fig. 3, B to D; fig. S7, B and C; and movie S12). This dissociation was transient, and by AnaIV, lymphatics resumed their niche connection.

Given the well-established role of lymphatic vessels in controlling tissue fluid balance and macromolecule efflux (*18*, *24*), we also assessed their drainage during hair regeneration. After intradermal injections, both Evans blue dye clearance (*25*) and OVA-Ax488 drainage into brachial lymph nodes were significantly delayed during the narrow hair cycle window when lymphatics were dissociated (Fig. 3E). Morphologically, the dissociated capillaries displayed signs of reduced permeability (*26*, *27*), with no major hair cycle–associated changes in blood vessel permeability (fig. S7, D and E). When fluid volume was artificially overloaded in skin, HFSCs precociously proliferated (fig. S7F), which is consistent with a role for transient lymphatic dilation in HF regeneration.

Although we cannot exclude the possibility that mechanical forces imposed by HF regeneration might influence SC–lymphatic dynamics, HFs were still growing downward after lymphatic–SC niche connections were restored. However, this timing did coincide with the return of bulge SCs to quiescence (*4*). Based on these considerations, lymphatic–SC connections seemed to be affecting HFSC behavior, but by mechanisms beyond HF downgrowth. Although outside the scope of this work, possibilities include draining proliferation stimuli (growth factors, metabolites, toxins) from the SC niche or, alternatively, producing inhibitory factors that keep HFSC self-renewal in check.

A lymphovascular switch at the onset of SC activation

When transplanted in vivo, cultured HFSCs can establish cycling HFs (*1*), suggesting that SCs participate in organizing their niche. This led us to speculate that HFSCs might be orchestrating lymphatic capillary connections. To address this possibility, we began by purifying and transcriptome-profiling HFSCs during (i) telogen, when SCs are quiescent and lymphatics are associated; and (ii) AnaII–III, when SCs are proliferative and lymphatics are dissociated (fig. S8A).

Established proangiogenic and lymphangiogenic factors, such as *Vegfa*, *Vegfb*, and *Vegfc*, showed little or no expression in either quiescent or proliferative bulge SCs and hence were unlikely to control bulge–lymphatic dynamics (Fig. 4A). Changes in stromal VEGFC expression and/or processing were also not detected during this time (fig. S8, B and C). Moreover, although a VEGFC gradient is known to elicit robust lymphatic VEGFR3-mediated sprouting, overt signs of enhanced sprouting were absent (fig. S7). Thus, SC–lymphatic dynamics appeared to be controlled by other factors.

A number of putative vascular regulators displayed expression patterns paralleling these dynamics. Most notably, *Angptl7* was expressed in telogen bulge SCs, whereas *Angptl4* and netrin-4 (*Ntn4*) were induced in AnaII–III bulge SCs concomitant with *Angptl7* down-regulation (Fig. 4A). These factors appeared to be bulge-specific (Fig. 4B).

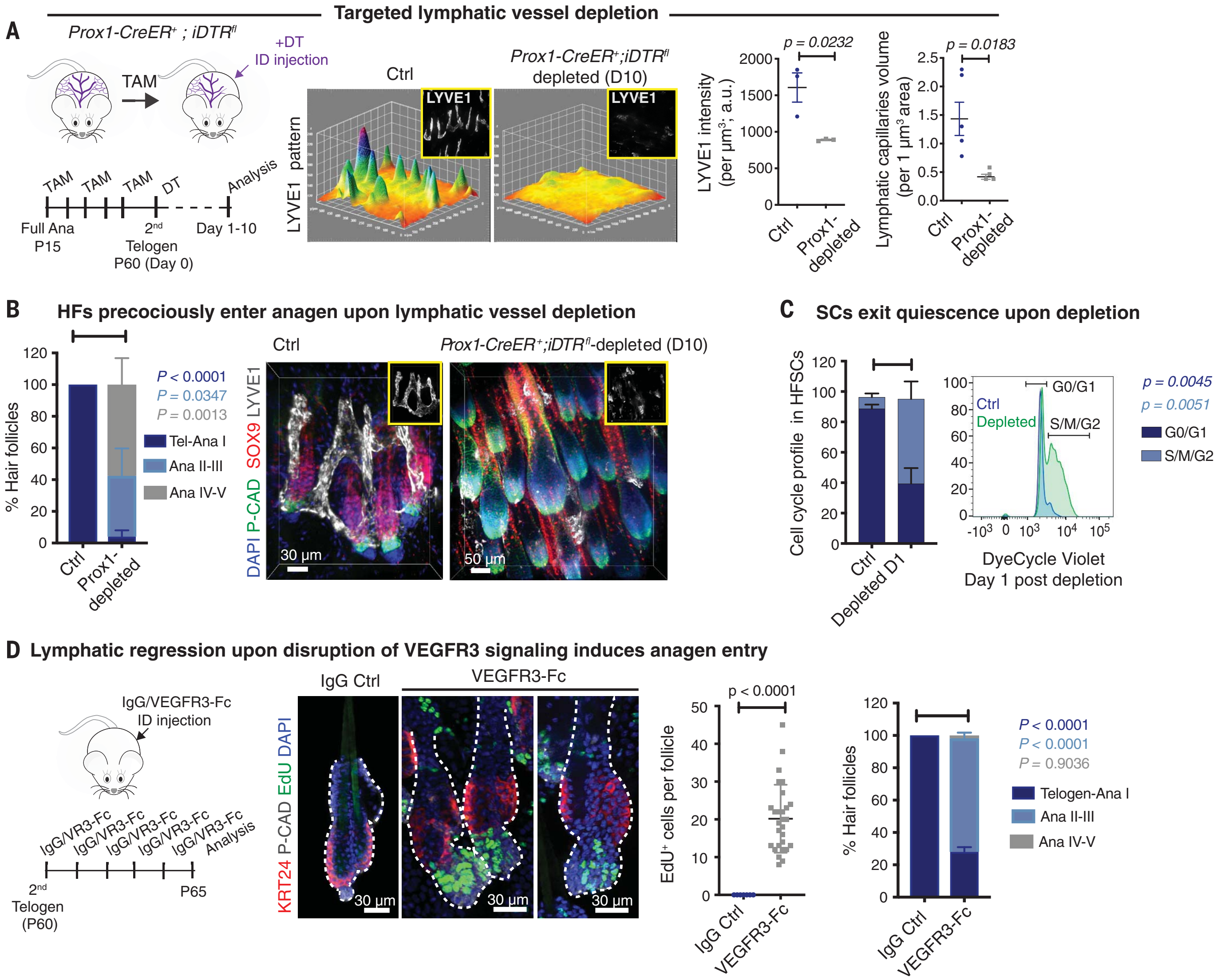


Fig. 2. Disrupting lymphatic capillaries triggers hair regeneration. (**A**) Timeline of experimental model and diphtheria toxin (DT)–induced depletion of lymphatic vessels during the long quiescence of second telogen. Quantifications: lymphatic intensity (n = 3); volume (n = 5 and 4 for Ctrl and *Prox1*-depleted, respectively; two-tailed unpaired t test). ID, intradermal; TAM, tamoxifen; P, postnatal day; Ctrl, control; D, day; a.u., arbitrary units. (**B**) Precocious anagen induction after lymphatic depletion, as demonstrated by HF growth and association with LYVE1$^+$ capillaries (n = 5, two-way ANOVA with Sidak's multiple comparisons test). P-CAD, P-cadherin. (**C**) Flow cytometry of HFSC cell cycle after lymphatic depletion (n = 3, two-way ANOVA with Sidak's multiple comparisons test). (**D**) Experimental strategy using decoy VEGF3-Fc to interrupt VEGFR3 signaling in dermal lymphatic vessels. EdU incorporation reveals anagen entry after treatment (n = 4, multiple EdU measurements, two-tailed unpaired t test). IgG, immunoglobulin G.

Focusing first on *Angptl7*, we find that prior RNA sequencing (RNA-seq) data (*28*) showed that *Angptl7* was not even expressed in the primed HG cells of telogen or early anagen HFs (Fig. 4C and fig. S8D). To delineate the detailed temporal changes of *Angptl7* within bulge SCs, we performed stage-specific single-cell RNA-seq. Three distinct temporal patterns emerged from t-distributed stochastic neighbor embedding (t-SNE) and unsupervised hierarchical clustering (Fig. 4D and fig. S8E). Using machine learning–based cell-cycle allocation analysis (*29*), it was concluded that proliferative cells resided in AnaII SCs, in agreement with our 5-ethynyl-2′-deoxyuridine (EdU) labeling studies. Additionally, in contrast to SC marker *Sox9*, which was maintained in bulge SCs at all stages, *Angptl7* transcription in bulge SCs diminished shortly after anagen onset, even sooner than that of established quiescence regulator *Nfatc1* (*28*). Moreover, the *Angptl7* expression dip was transient, occurring only between AnaII and AnaIII (Fig. 4E).

Angptl7 dynamics were recapitulated at the chromatin level. As judged by chromatin immunoprecipitation sequencing (ChIP-seq) analyses of fluorescence-activated cell sorting (FACS)–purified HF cells (*30*), the *Angptl7* chromatin state was active in telogen (H3K27ac$^+$), poised in AnaIII (H3K27ac^{neg}H3K4me1$^+$), and silent in AnaVI differentiated cells (H3K27ac^{neg}H3K4me1^{neg}) (fig. S9A). Moreover, when either of the two *Angptl7* locus–accessible chromatin elements (*30*) were used as enhancers to drive enhanced green fluorescent protein (*eGFP*) expression in vivo, reporters faithfully recapitulated *Angptl7*'s temporal dynamics in bulge expression during the hair cycle (Fig. 4F and fig. S9B). These results intimated that changes in SC-derived *Angptl7* transcription are involved in reshaping the regenerative microenvironment of the niche.

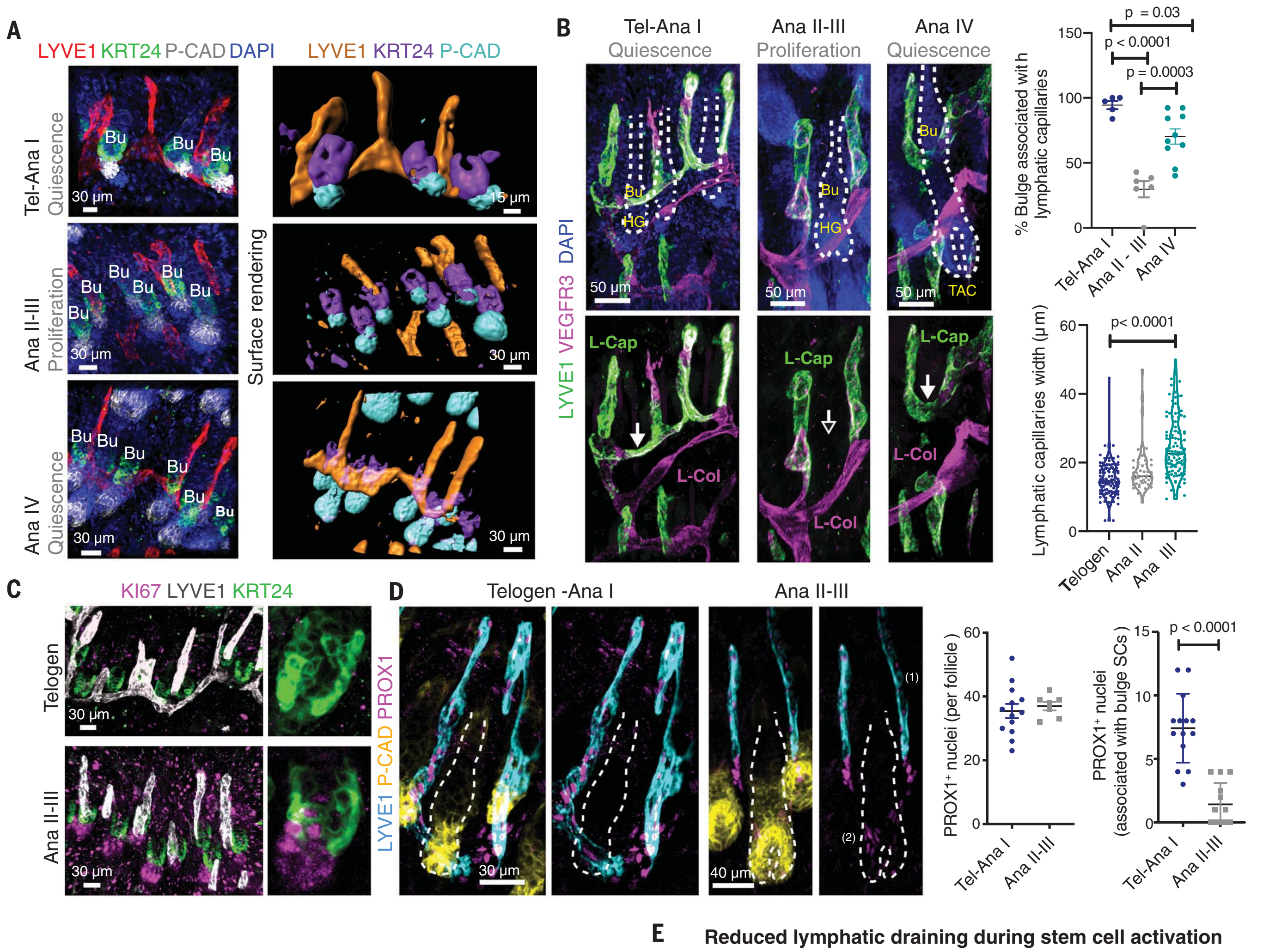


Fig. 3. HFSC–lymphatic associations are dynamic. (**A**) 3D imaging and surface rendering of cleared skin tissue, demonstrating LYVE1$^+$ lymphatic capillary dissociation from KRT24$^+$ bulge SCs during AnaII–III of the hair cycle. P-CAD^{high} cells are HFSC progeny. (**B**) Despite dissociating (empty arrow for dissociation versus white arrow for association) from the SC niche, lymphatic capillaries remain connected to their underlying collecting vessels (LYVE1^{neg}VEGFR3$^+$) during AnaII–III. Quantifications of temporal changes in SC–capillary associations and lymphatic width (n = 5, 6 and n = 5, one-way ANOVA; Tukey's multiple comparisons test). TAC, transient amplifying cells. (**C**) Capillary dissociation correlates with bulge SC self-renewal (Ki67$^+$). (**D**) PROX1$^+$ lymphatics dissociate from the bulge during AnaII–III but do not disappear (1 indicates Prox1$^+$capillaries, 2 indicates Prox1$^+$collecting; n = 3; two-tailed unpaired t test). (**E**) Reduced lymphatic drainage capacity in AnaII–III, as measured by intradermal injections of Evans Blue (EB; n = 3, one-way ANOVA; Tukey's multiple comparisons test) and OVA-Ax488 efflux to brachial lymph nodes (LN; n = 3 and 6 for Tel–AnaI and AnaII–III, respectively; two-tailed unpaired t test). OD, optical density.

A SC-driven secretome switch is required for lymphatic remodeling

Angiopoietin-like (ANGPTL) proteins do not bind to classical angiopoietin receptors, and knowledge of these orphan ligands is still scant (*31*, *32*). To address whether *Angptl7* down-regulation functions in HFSC activation, we engineered mice to harbor an *Angptl7* transgene that is selectively doxycycline-inducible in skin progenitors (*33*). We then induced at telogen and examined the consequences of maintaining ANGPTL7 through early stages of the newly entered hair cycle (fig. S10, A to E).

In contrast to the lymphatic–SC niche remodeling of normal HFs, lymphatic capillaries remained tightly associated and the bulge remained quiescent when *Angptl7* was sustained (Fig. 5A). Consistent with a nonautonomous role for ANGPTL7 in controlling stemness activity, recombinant ANGPTL7 did not appreciably alter colony formation of HFSCs in vitro. Rather, concomitant with sustained lymphatic efflux and drainage capacity, entry into the hair cycle was markedly delayed (Fig. 5B and fig. S10, C to E). Because a role for ANGPTL7 in angiogenesis had been suggested, we looked for, but did not find, perturbations in blood vessel density and coverage in our *Angptl7*-induced skin (fig. S10E). ANGPTL7

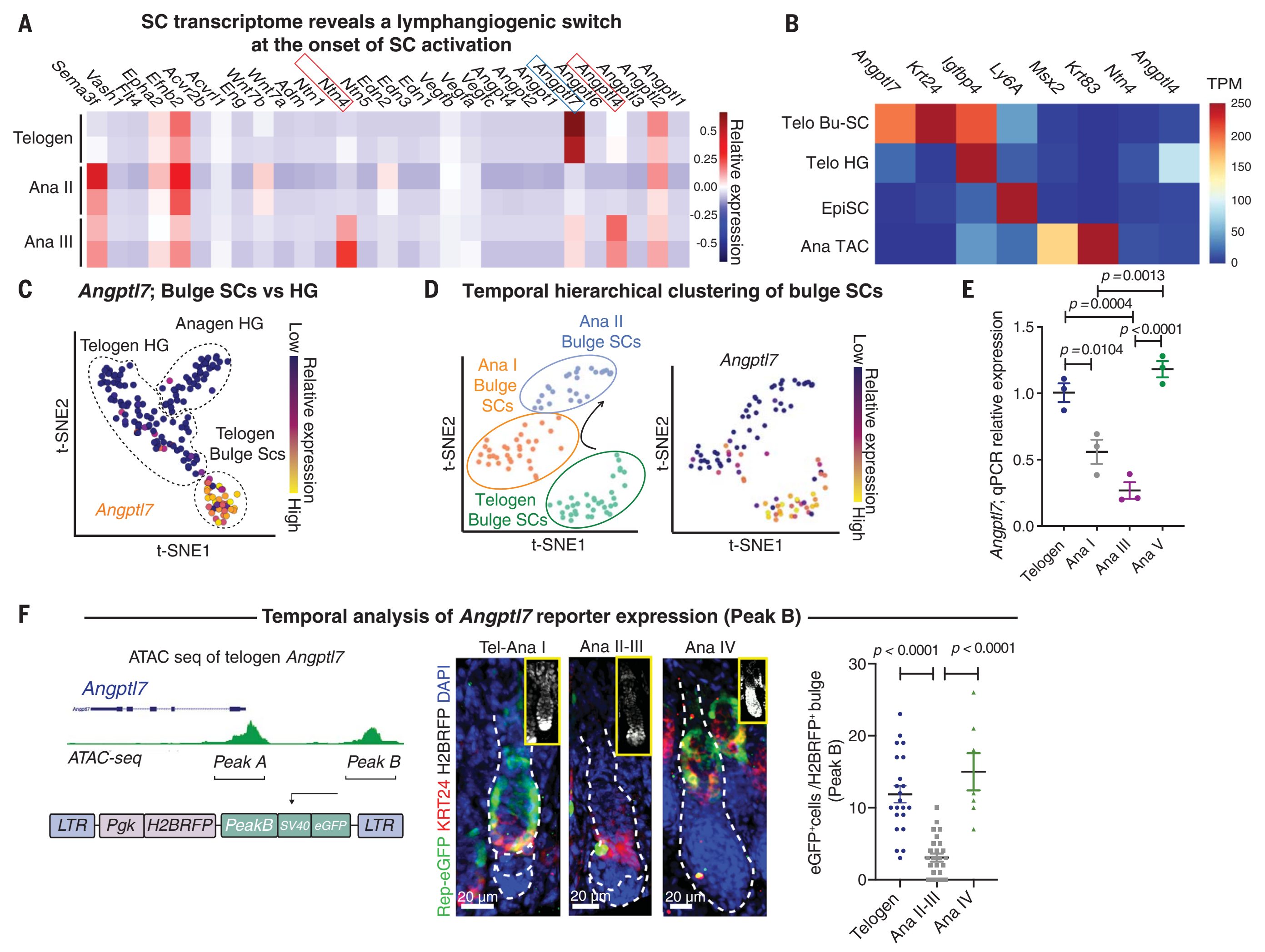


Fig. 4. HFSCs promote a lymphovascular switch at hair regeneration onset. (**A**) Heatmaps of FACS-purified bulge SCs from telogen, AnaII, and AnaIII HFs. Boxed transcripts are markedly down (blue) or up (red) upon HFSC activation. (**B**) Transcripts per million (TPM) heatmap showing paucity of *Angptl7*, *Angptl4*, and *Ntn4* in SCA1$^+$epidermal SCs (EpiSCs) or HFSC progeny (MSK2$^+$Krt83$^+$ TACs; IGBP4$^+$ KRT^{neg} HGs). (**C** and **D**) Single-cell transcriptome clustering of (C) telogen bulge SCs, telogen HG cells, and AnaI HG cells (*28*); and (D) telogen, AnaI, and AnaII bulge SCs. Each cell is represented as a dot. (**E**) Quantitative polymerase chain reaction (qPCR) validation of temporal *Angptl7* expression in bulge SCs (n = 3, one-way ANOVA; Tukey's multiple comparisons test). (**F**) Accessible *Angptl7* locus telogen bulge SC chromatin peaks (left) were used to drive *eGFP* in mice. Peak B *eGFP* dynamics (right) mirror temporal *Angptl7* transcriptome changes (n = 5, 5, and 3 mice analyzed for Telogen, AnaII–III, and AnaIV, respectively; one-way ANOVA; Tukey's multiple comparisons test). ATAC seq, assay for transposase-accessible chromatin using sequencing; LTR, long terminal repeats; Pgk, phosphoglycerate kinase promoter; H2BRFP, histone H2B red fluorescent protein; SV40, minimal simian virus 40 promoter; Rep-eGFP, reporter-enhanced green fluorescent protein.

overexpression also did not affect levels of CCBE1, a protein required for VEGFC-mediated lymphangiogenesis (fig. S10F).

Because the hair cycle was delayed upon sustaining *Angptl7*, we wondered whether ANGPTL7 dynamics might be important for other regenerative responses. We focused on hair plucking (waxing), as it mechanically perturbs the SC niche, generating a wound-like response (*3*, *34*). Plucking not only activated hair regeneration, as expected, it also restricted lymphatic drainage and disrupted lymphatic–SC associations (fig. S11, A and B). Moreover, when *Angptl7* was sustained, wound-induced hair regeneration was abrogated; conversely, after skin injury in a normal setting, *Angptl7* gene expression plummeted (fig. S11, C and D).

Turning to the flip side of the secretome switch, we engineered doxycycline-inducible versions of *Angptl4*. When induced in telogen, ANGPTL4 precociously disrupted lymphatic capillaries surrounding the bulge, reducing lymphatic drainage and activating HFSCs (Fig. 5, C and D, and fig. S12). Despite a purported role for ANGPTL4 in blood vasculature (*35*), endothelial vessel density was largely unchanged, both in AnaII–III of the normal hair cycle and in telogen of the *Angptl4*-induced hair cycle. Similar results were obtained with NTN4, previously implicated in endothelial cell biology (*36*–*38*) (fig. S12).

Our data suggested that the HFSC-derived secretome can influence lymphatic dynamics directly, which we tested by evaluating lymphatic tube formation in a 3D-Matrigel system. Formation of tubelike structures was enhanced with ANGPTL7 and impaired with ANGPTL4, supporting this hypothesis. Consistent with a nonautonomous role, SC-driven factors in vitro neither affected HFSC growth nor stimulated immune cell migration (fig. S13).

Although ANGPTL and NTN receptors are poorly studied, transcriptional landscaping of freshly isolated endothelial cells from skin

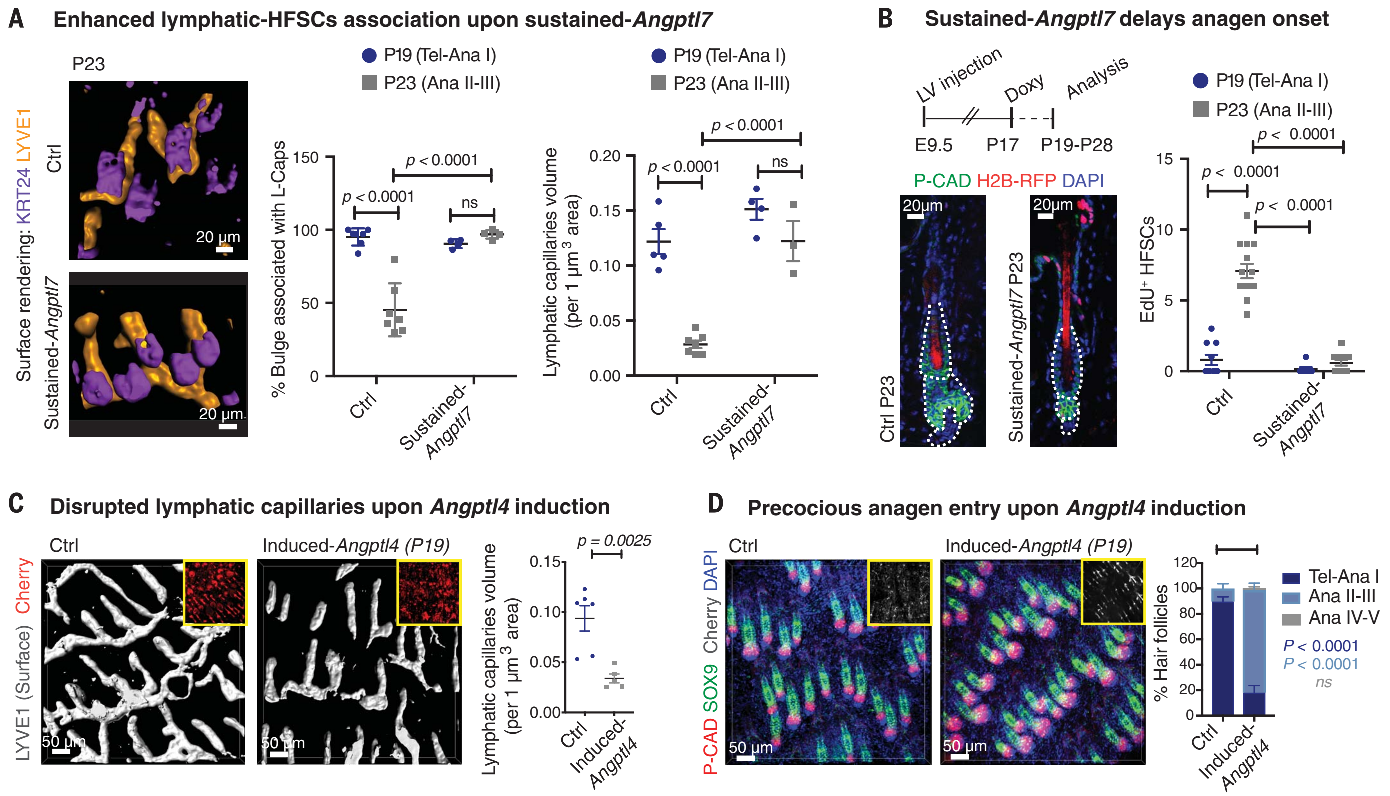


Fig. 5. SC secretome switch promotes lymphatic remodeling. (**A**) (Left) Lymphatic–bulge SC connections and quantifications (n = 7 $rtTA^{neg}$ control; n = 4 $rtTA^{+}$ sustained-*Angptl7*). (Right) Capillary volumes ($n \geq 4$ per condition per genotype; two-way ANOVA; Tukey's multiple comparisons test). ns, not significant. (**B**) Doxycycline-induced *Angptl7* delays anagen onset and bulge-SC proliferation (n = 4 per condition per genotype; two-way ANOVA; Tukey's multiple comparisons test). E, embryonic day. (**C** and **D**) Induced *Angptl4* in telogen causes reduced lymphatic capillary density ($n \geq 5$ per condition per genotype, two-tailed unpaired t test) and precocious anagen entry ($n \geq 3$ per condition per genotype, two-way ANOVA with Sidak's multiple comparisons test).

revealed preferential expression of their putative receptors by the lymphatic vessels (fig. S14A). Telogen-associated lymphatic capillaries also preferentially expressed WNT inhibitors. As WNT signaling is critical for HFSC activation and hair cycling (*30*, *39*), these correlations are suggestive of an additional layer of regulation by which lymphatic capillaries might coordinate SC regeneration.

Single-cell analyses of lymphatic capillary cells isolated and characterized from telogen and from anagen skins revealed similar transcriptomic patterns, with only 4 to 5% of mRNAs changed temporally by >2× (fig. S14B). Of these genes, the most notable changes appeared to be in lymphatic tube formation and fluid dynamics. Thus, if capillary growth factors participate in maintaining HFSC quiescence, their significance likely resides in the dynamic regulation of lymphatic–SC connections rather than their differential transcription during the hair cycle.

A role for the lymphatic network in integrating SC-niche behavior across a tissue

Turning to the physiological relevance of the SC–lymphatic connection, we used our powerful in utero lentiviral delivery method to selectively and efficiently knock down *Angptl7* in skin progenitors with one of two different short hairpin RNAs (shRNAs). When *Scramble* controls were in telogen, *shAngptl7* HFs had already entered anagen. Full anagen follicles were interspersed with telogen ones, indicating hair cycle asynchrony (Fig. 6A and fig. S15A).

The failure of HFSCs to generate ANGPTL7 profoundly affected lymphatic biology. Lymphatics were discontinuous and abnormally dilated, and they displayed impaired drainage, as judged both functionally and morphologically (Fig. 6B and fig. S15B). This chronic lymphatic dysfunction was associated with HF hyperplasia and reduced bone morphogenetic protein (BMP)/pSMAD1/5/9 signaling, which are essential for maintaining SC quiescence (*3*) (Fig. 6, B and C; fig. S15C; and movie S13). Hyperplastic HFs were associated with the most highly dilated lymphatic capillaries, which maintained their identity but displayed reduced drainage (Fig. 6, B and D; fig. S15D; and movie S14). Blood vessel density was also increased at hyperplastic HFs (fig. S15E and movie S15), although this was unlikely to have driven lymphatic remodeling, given that SCs associated primarily with lymphatic capillaries in normal homeostasis.

The asynchrony across the hair coat did not appear to stem from variations in transgene integration, as this would have generated clonal patches of HFs at specific cycle stages, which we did not see. However, to unequivocally demonstrate that HF asynchrony arose from perturbations in the SC–lymphatic crosstalk, and not our lentiviral delivery method, we used $Flt4^{Chy}$ mice, which harbor a mutant *Vegfr3* allele, creating VEGFR3 dimers with dysfunctional tyrosine kinase activity and dysfunctional lymphatics (*40*). $Flt4^{Chy}$ mice recapitulated the asynchrony of bulge-SC niches, accompanied with dysfunctional lymphatic vessels (Fig. 6E and fig. S15, F and G). Taken together, these data underscore the importance of lymphatic capillary dynamics, driven by SCs, in integrating SC–niche behavior across a tissue.

Lymphatic capillaries as a dynamic SC-niche newcomer that coordinates SC behavior

The niche microenvironments of quiescent SCs, such as those of bone marrow, muscle, and HFs, provide the input signals that keep these SCs in an undifferentiated, inactive state (*41*). Niche–SC interactions must be dynamic in order to mobilize SCs to regenerate tissues. Additionally, SCs have an intrinsic ability to

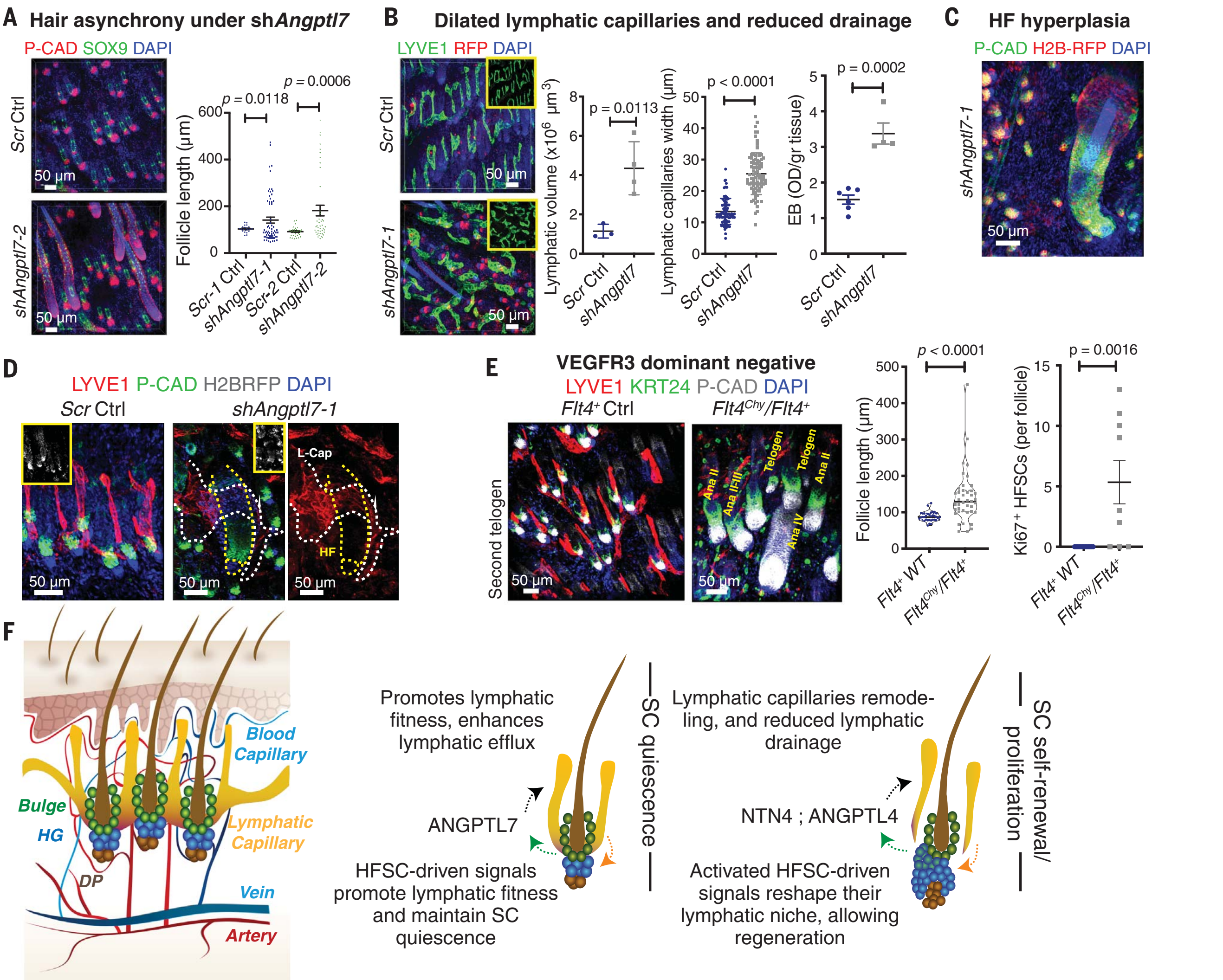


Fig. 6. Secretome switch integrates lymphatic–SC niches across the tissue. (**A**) *shAngptl7* knockdown results in HF asynchrony by P19 (n = 4 per condition per shRNA, one-way ANOVA with Dunnett's multiple comparisons test). (**B**) *shAngptl7* in HFSCs causes lymphatic capillary dilation ($n \geq 3$ per shRNA, two-tailed unpaired t test). Quantifications of lymphatic drainage capacity are shown at right, as measured by Evans blue efflux ($n \geq 4$ per shRNA per measurement, two-tailed unpaired t test). (**C**) HF hyperplasia after *Angptl7* knockdown and (**D**) hyperplastic HF association with the most highly dilated lymphatic capillaries. (**E**) Expressing a dominant-negative VEGFR3, $Flt4^{Chy}/Flt4^{+}$ mice exhibit hair-cycling asynchrony and increased proliferation at P65, when control HFs are in second telogen (n = 3, two-tailed unpaired t test). (**F**) Model of the HFSC secretome switch.

communicate with their neighbors and reshape their niche (*34*).

Our study identifies lymphatic capillaries as dynamic SC-niche elements that integrate SC niches and synchronize SC behavior across the hair coat. We have learned that, to mobilize HFSCs for either normal regenerative demands or those induced by injury, these SCs undergo a secretome switch that triggers transient dissociation of lymphatics from the niche (Fig. 6F).

Although the possible ways by which lymphatics control HFSCs are numerous, and the communication circuits are likely complex, our genetic manipulations of the lymphatic–SC niche connection underscore its functional importance in balancing SC self-renewal and quiescence. Indeed, as soon as bulge SCs launch production of their committed proliferative progeny and self-renew, the lymphatic capillaries resume connections with their SC neighbors, and niche quiescence is restored. The process is a two-way street, because the dynamic remodeling of lymphatic capillaries is orchestrated by bulge SCs, whereas the lymphatics govern SC behavior.

Drainage of tissue interstitial fluids and macromolecules has been a subject of interest for decades. We have unearthed a need to balance cutaneous influx of fluids and macromolecules in controlling SC behavior. Future studies will be needed to dissect the impact of interstitial fluid composition, extracellular macromolecule dynamics, and immune cell efflux on stem cell biology. Additionally, short remodeling duration suggests that lymphatic dissociation from the SC niche may be deleterious, perhaps rendering SCs transiently vulnerable to toxins or increased fluid pressure.

Our discovery that lymphatics localize to both mouse and human HFSC niches suggests that the need to establish such connections is not only physiologically important but also evolutionarily conserved, raising the possibility that lymphatic capillaries may participate in other SC niches to meet their specialized regenerative demands. With our newfound understanding of skin SC–lymphatic interactions, it now merits addressing whether the stem cell exhaustion that accompanies wound-healing defects and hair loss in aging and in

patients with lymphedema (*24*, *42*) might be rooted in decline of lymphatics and interstitial fluid draining. If such links exist, targeting lymphatic function could prove to be a promising preventative therapeutic target for hair loss and wound repair.

REFERENCES AND NOTES

1. K. A. U. Gonzales, E. Fuchs, *Dev. Cell* **43**, 387–401 (2017).
2. L. Li, H. Clevers, *Science* **327**, 542–545 (2010).
3. Y. C. Hsu, L. Li, E. Fuchs, *Nat. Med.* **20**, 847–856 (2014).
4. S. J. Morrison, A. C. Spradling, *Cell* **132**, 598–611 (2008).
5. Y. Ge, E. Fuchs, *Nat. Rev. Genet.* **19**, 311–325 (2018).
6. M. V. Plikus *et al*., *Nature* **451**, 340–344 (2008).
7. E. Festa *et al*., *Cell* **146**, 761–771 (2011).
8. J. Goldstein *et al*., *Genes Dev.* **28**, 983–994 (2014).
9. A. Durward, K. M. Rudall, in *The Biology of Human Hair Growth* (Academic Press, 1958), pp. 189–218.
10. O. J. Tamplin *et al*., *Cell* **160**, 241–252 (2015).
11. M. J. Kiel *et al*., *Cell* **121**, 1109–1121 (2005).
12. Y. Kunisaki *et al*., *Nature* **502**, 637–643 (2013).
13. M. Tavazoie *et al*., *Cell Stem Cell* **3**, 279–288 (2008).
14. W. Tang *et al*., *Science* **322**, 583–586 (2008).
15. S. Yoshida, M. Sukeno, Y. Nabeshima, *Science* **317**, 1722–1726 (2007).
16. D. Peña-Jimenez *et al*., *EMBO J.* **38**, e101688 (2019).
17. D. S. S. Richardson, J. W. W. Lichtman, *Cell* **162**, 246–257 (2015).
18. K. Alitalo, T. Tammela, T. V. Petrova, *Nature* **438**, 946–953 (2005).
19. J. A. Nowak, L. Polak, H. A. Pasolli, E. Fuchs, *Cell Stem Cell* **3**, 33–43 (2008).
20. R. S. Srinivasan *et al*., *Genes Dev.* **21**, 2422–2432 (2007).
21. T. Buch *et al*., *Nat. Methods* **2**, 419–426 (2005).
22. N. Ali *et al*., *Cell* **169**, 1119–1129.e11 (2017).
23. T. Mäkinen *et al*., *Nat. Med.* **7**, 199–205 (2001).
24. S. Da Mesquita *et al*., *Nature* **560**, 185–191 (2018).
25. R. Huggenberger *et al*., *Blood* **117**, 4667–4678 (2011).
26. P. Baluk *et al*., *J. Exp. Med.* **204**, 2349–2362 (2007).
27. F. Zhang *et al*., *Science* **361**, 599–603 (2018).
28. H. Yang, R. C. Adam, Y. Ge, Z. L. Hua, E. Fuchs, *Cell* **169**, 483–496.e13 (2017).
29. A. Scialdone *et al*., *Methods* **85**, 54–61 (2015).
30. R. C. Adam *et al*., *Cell Stem Cell* **22**, 398–413.e7 (2018).
31. C. C. Zhang *et al*., *Nat. Med.* **12**, 240–245 (2006).
32. J. Zheng *et al*., *Nature* **485**, 656–660 (2012).
33. S. Beronja, G. Livshits, S. Williams, E. Fuchs, *Nat. Med.* **16**, 821–827 (2010).
34. C. Blanpain, W. E. Lowry, A. Geoghegan, L. Polak, E. Fuchs, *Cell* **118**, 635–648 (2004).
35. F. Bäckhed, P. A. Crawford, D. O'Donnell, J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 606–611 (2007).
36. X. Lu *et al*., *Nature* **432**, 179–186 (2004).
37. B. D. Wilson *et al*., *Science* **313**, 640–644 (2006).
38. V. Cirulli, M. Yebra, *Nat. Rev. Mol. Cell Biol.* **8**, 296–306 (2007).
39. W. H. Lien *et al*., *Nat. Cell Biol.* **16**, 179–190 (2014).
40. M. J. Karkkainen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12677–12682 (2001).
41. S. J. Morrison, D. T. Scadden, *Nature* **505**, 327–334 (2014).
42. S. Karaman *et al*., *Angiogenesis* **18**, 489–498 (2015).

ACKNOWLEDGMENTS

We thank The Rockefeller University's FACS facility (S. Mazel); Rockefeller's and Weill Cornell's high-throughput sequencing facilities; K. Minkis (Weill Cornell) for human biopsies; and G. Oliver (Northwestern), M. Gunzer (Univ. of Duisburg-Essen), G. Zarkada (Yale), and P. Rajasethupathy, P. Cohen, S. Shaham, and G. D. Victora (Rockefeller) for discussions. We thank the Fuchs laboratory for technical assistance (N. Gomez, I. Matos, K. Lay, B. Hurwitz, M. ElJaby, E. Wong, and M. Nikolova); mouse handling and experiments (M. Sribour, L. Hidalgo, and L. Polak); anti-KRT24 antibody (E. Wong); and discussions (S. Liu, F. Garcia Quiroz, S. Naik, Y. Ge, M. Laurin, S. Ellis, V. Fiore, K. Gonzales, R. C. Adam, and R. E. Niec). **Funding:** S.G.-C. received postdoctoral support from the Human Frontier Science Program (LT001519/2017), EMBO (ALTF 1239-2016), and the Revson Foundation. E.F. is an HHMI investigator. S.C.B. received predoctoral support from the National Cancer Institute (F31CA236465) and the Tri-Institutional Medical Scientist Training Program (T32GM007739). This work was supported by NIH grants R01-AR050452 and R01-AR31737 to E.F. **Author contributions:** S.G.-C. and E.F. conceptualized the study, designed the experiments, and wrote the manuscript. H.Y. helped with cloning strategies, analyzed sequencing data, and contributed to lymphatic-depletion experiments. S.C.B performed and analyzed in vitro experiments. Y.M. performed and analyzed immune profiling. J.L. performed in utero lentiviral injections. R.P.K., X.L., and B.J.M. helped with mouse models and contributed to lymphatic-related in vitro assays. J.C.-R. helped with skin sections. S.G.-C. performed all other experiments. E.F. supervised the project. **Competing interests:** B.J.M. is an inventor on patent application number 62112273, submitted by Memorial Sloan Kettering Cancer Center, that covers the treatment of lymphedema with topical formulations. **Data and materials availability:** All materials are available by material transfer agreements. Newly generated data have been deposited in NCBI's GenBank, and their accession numbers are listed in table S3.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1218/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References (*43*–*54*)
Movies S1 to S15

View/request a protocol for this paper from *Bio-protocol*.

19 June 2019; accepted 15 October 2019
Published online 31 October 2019
10.1126/science.aay4509

QUANTUM INFORMATION

Electrical and optical control of single spins integrated in scalable semiconductor devices

Christopher P. Anderson[1,2]*, Alexandre Bourassa[1]*, Kevin C. Miao[1], Gary Wolfowicz[1], Peter J. Mintun[1], Alexander L. Crook[1,2], Hiroshi Abe[3], Jawad Ul Hassan[4], Nguyen T. Son[4], Takeshi Ohshima[3], David D. Awschalom[1,2,5]†

Spin defects in silicon carbide have the advantage of exceptional electron spin coherence combined with a near-infrared spin-photon interface, all in a material amenable to modern semiconductor fabrication. Leveraging these advantages, we integrated highly coherent single neutral divacancy spins in commercially available p-i-n structures and fabricated diodes to modulate the local electrical environment of the defects. These devices enable deterministic charge-state control and broad Stark-shift tuning exceeding 850 gigahertz. We show that charge depletion results in a narrowing of the optical linewidths by more than 50-fold, approaching the lifetime limit. These results demonstrate a method for mitigating the ubiquitous problem of spectral diffusion in solid-state emitters by engineering the electrical environment while using classical semiconductor devices to control scalable, spin-based quantum systems.

Solid-state defects have enabled many proof-of-principle quantum technologies in quantum sensing (*1*), computation (*2*), and communications (*3*). These defects exhibit atom-like transitions that have been used to generate spin-photon entanglement and high-fidelity single-shot readout (*4*), enabling demonstrations of long-distance quantum teleportation, entanglement distillation, and loophole-free tests of Bell's inequalities (*3*). However, fluctuating electric fields and uncontrolled charge dynamics have limited many of these technologies (*1*, *4*–*7*). For example, lack of charge stability and of photon indistinguishability are major problems that reduce entanglement rates and fidelities in quantum communication experiments (*4*–*6*). In particular, indistinguishable and spectrally narrow photon emission is required to achieve high-contrast Hong-Ou-Mandel interference (*8*). This indistinguishability has been achieved with some quantum emitters through dc Stark tuning the optical lines into mutual resonance (*9*, *10*). A variety of strategies (*1*, *6*, *11*–*13*) have also been proposed to reduce spectral diffusion (*14*) and blinking (*15*), but consistently achieving narrow and photostable spectral lines remains an outstanding challenge (*16*). In addition, studies of charge dynamics (*17*, *18*) have enabled quantum-sensing improvements (*1*, *7*) and spin-to-charge conversion (*19*), allowing electrical readout of single-spin defects (*20*). However, these experiments have largely been realized in materials such as diamond, in which scalable nanofabrication and doping techniques are difficult to achieve.

By contrast, the neutral divacancy (VV^0) defect in silicon carbide (SiC) presents itself as a candidate spin qubit in a technologically mature host, allowing for flexible fabrication, doping control, and availability on the wafer scale. These defects display many attractive properties, including all-optical spin initialization and readout (*21*), long coherence times (*22*), nuclear spin control (*23*), as well as a near-infrared high-fidelity spin-photon interface (*24*). However, VV^0 has displayed relatively broad optical lines (*24*), charge instability (*18*), and relatively small Stark shifts (*10*). Furthermore, the promise of integration into classical semiconducting devices remains largely unexplored.

Here, we use the mature semiconductor technology that SiC provides to create a p-i-n structure that allows tuning of the electric field and charge environment of the defect. First, we isolate and perform high-fidelity control on highly coherent single spins in the device. We then show that these devices enable wide dc Stark tuning while maintaining defect symmetry. We also demonstrate that charge depletion in the device mitigates spectral diffusion, thus greatly narrowing the linewidths in the optical fine structure. Finally, we use this device as a testbed to study the photoionization dynamics of single VV^0, resulting in a method for deterministic optical control of the defect charge state.

The effects presented here suggest that doped SiC structures are flexible and scalable quantum platforms hosting long-lived, single-spin qubits with an electrically tunable, high-quality optical interface. The demonstrated reduction in electric field noise can lead to increased spin coherence (*25*) and electrical tuning of "dark" spins in quantum sensing (*26*), whereas charge control could extend the memory time of nuclear spins (*27*). Additionally, this platform opens new avenues for spin-to-charge conversion, electrically driven single-photon emission (*28*), electrical control (*29*), and readout (*20*, *30*, *31*) of single spins in SiC CMOS (complementary metal oxide semiconductor)-compatible and optoelectronic semiconductor devices.

Isolated single defects in a semiconductor device

We first isolated and controlled single VV^0 in a 4H-SiC p-i-n diode created through commercial growth of doped SiC epilayers. After growth, we electron irradiated and annealed our samples to create single, isolated VV^0 defects. We fabricated microwave striplines and ohmic contact pads, allowing for spin manipulation and electrical gating (Fig. 1A) (*32*). In contrast to other defects in SiC, such as the isolated silicon vacancy (*33*), the divacancy is stable above 1600°C (*34*), making it compatible with device processing and high-temperature annealing to form ohmic contacts.

Spatial photoluminescence (PL) scans of the device showed isolated emitters corresponding to single VV^0 (Fig. 1B), as confirmed by second-order correlation ($g^{(2)}$) measurements (Fig. 1B, inset) (*32*). The location in depth of the observed defects is consistent with isolation to the i-type layer. This is to be expected because formation energy calculations (*35*) indicate that the neutral charge state is energetically favorable when the Fermi level is between ~1.1 and 2 eV, and this condition must be satisfied somewhere in the i-type layer (*32*, *36*). This depth localization provides an alternative to delta doping (*37*), which is not possible with intrinsic defects, facilitating positioning and control in fabricated devices (fig. S1). Additionally, owing to the diode's highly rectifying behavior at low temperature, large reverse biases are possible with low current (Fig. 1C) (*32*).

Sweeping the frequency of a narrow-line laser, we obtained photoluminescence excitation (PLE) spectra of the optical fine structure of these single defects (Fig. 1D). Using the observed transitions for resonant readout and preparation, we performed high-contrast Rabi oscillations of isolated VV^0 in the p-i-n structure (Fig. 1E) (*32*). The contrast exceeded 98%, improving on previous demonstrations through the use of resonant spin polarization (*24*). Additionally, a single-spin Hahn-echo decay time of 1.0 ± 0.1 ms was measured for spins in the device (Fig. 1F), consistent with previous ensemble measurements (*22*). The long Hahn-echo times and high-fidelity control demonstrate that integration into the semiconductor structures does not degrade the spin properties of VV^0. This isolation and control of highly coherent spin qubits achieved

[1]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA. [2]Department of Physics, University of Chicago, Chicago, IL 60637, USA. [3]National Institutes for Quantum and Radiological Science and Technology, 1233 Watanuki, Takasaki, Gunma 370-1292, Japan. [4]Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden. [5]Center for Molecular Engineering and Materials Science Division, Argonne National Laboratory, Lemont, IL 60439, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: awsch@uchicago.edu

in these functioning semiconductor devices unlocks the potential for integration with a wide range of classical electronic technologies.

Large Stark shifts in a p-i-n diode

Because the (*hh*) and (*kk*) divacancies (*32*) in SiC are nominally symmetric along the c-axis (growth axis), the geometry of the diode allows for large electric fields that mostly conserve the symmetry of the defect. Therefore, wide tuning of the VV^0 optical structure is possible while reducing unwanted mixing from transverse or symmetry-breaking components of the excited-state Hamiltonian (*9*, *24*, *38*). Because the i-type region can be relatively thin (10 μm here), the applied voltage is dropped over a much smaller region than if a bulk sample were used (*10*), leading to significantly larger Stark shifts for a given applied voltage. In principle, this region can be reduced to a thickness that exceeds limitations from optical access with metal planar gates (limited by the optical spot size of ~1 μm). Furthermore, it is possible to use doped layers as in situ transparent native contacts to Stark tune and control localized defects in suspended photonic or phononic structures (*39*), enabling complex hybrid electrical, photonic, and phononic devices.

In our p-i-n junction device, we applied up to 420 V in reverse bias. Our results show Stark tuning of several hundreds of gigahertz on different defects of the same type and on inequivalent lattice sites, where the Stark shift was between 0.4 and 3.5 GHz/V after a threshold was passed (Fig. 2A). For example, we observed a (*hh*) divacancy shifted by >850 GHz (2.5 meV) at a reverse bias of 420 V and a (*kh*) divacancy shifted by >760 GHz at a reverse bias of 210 V (Fig. 2B). These shifts are among the largest reported for any single-spin defect to date and were only limited by the voltage output of our source. We expect that, owing to the high dielectric breakdown field of SiC, even higher shifts of a few terahertz are possible (*32*). The high-field limit of these shifts corresponds to estimated dipole moments ($d_{||}$) of 11 GHz m/MV and 4.5 GHz m/MV for (*hh*) and (*kk*) divacancies, respectively, consistent with previous reports (*10*, *40*). For the (*kh*) basal divacancy observed, the estimated transverse dipole moment is $d_{\perp}$ ~ 35 GHz m/MV. Furthermore, because the Stark shift represents a measure of the local electric field, we conclude that negligible field is applied to the VV^0 before a certain threshold voltage where the depletion region reaches the defect (*41*). This results from nonuniform electric fields in the diode caused by residual n-type dopants in the intrinsic region [Fig. 2C (*32*)].

Overall, our system could be used as a widely frequency-tunable, spectrally narrow source of single photons. In particular, our system enables one of the highest Stark shift-to-linewidth ratios (>40,000) obtained in any solid-state single-photon source (table S1). These characteristics make this system ideally suited for tuning remote defects into mutual resonance and for frequency multiplexing of entanglement channels (*42*). The tunability range is so wide that it could even enable the tuning of a (*hh*) divacancy into resonance with a (*kk*) divacancy, allowing for interference and entanglement between different species of defects. This wide tunability stems from the rectification behavior of the diode, which allows large electric fields without driving appreciable currents that can degrade spin and optical properties. Furthermore, the observed sensitivity of the optical structure of single VV^0 defects could serve as a nanoscale

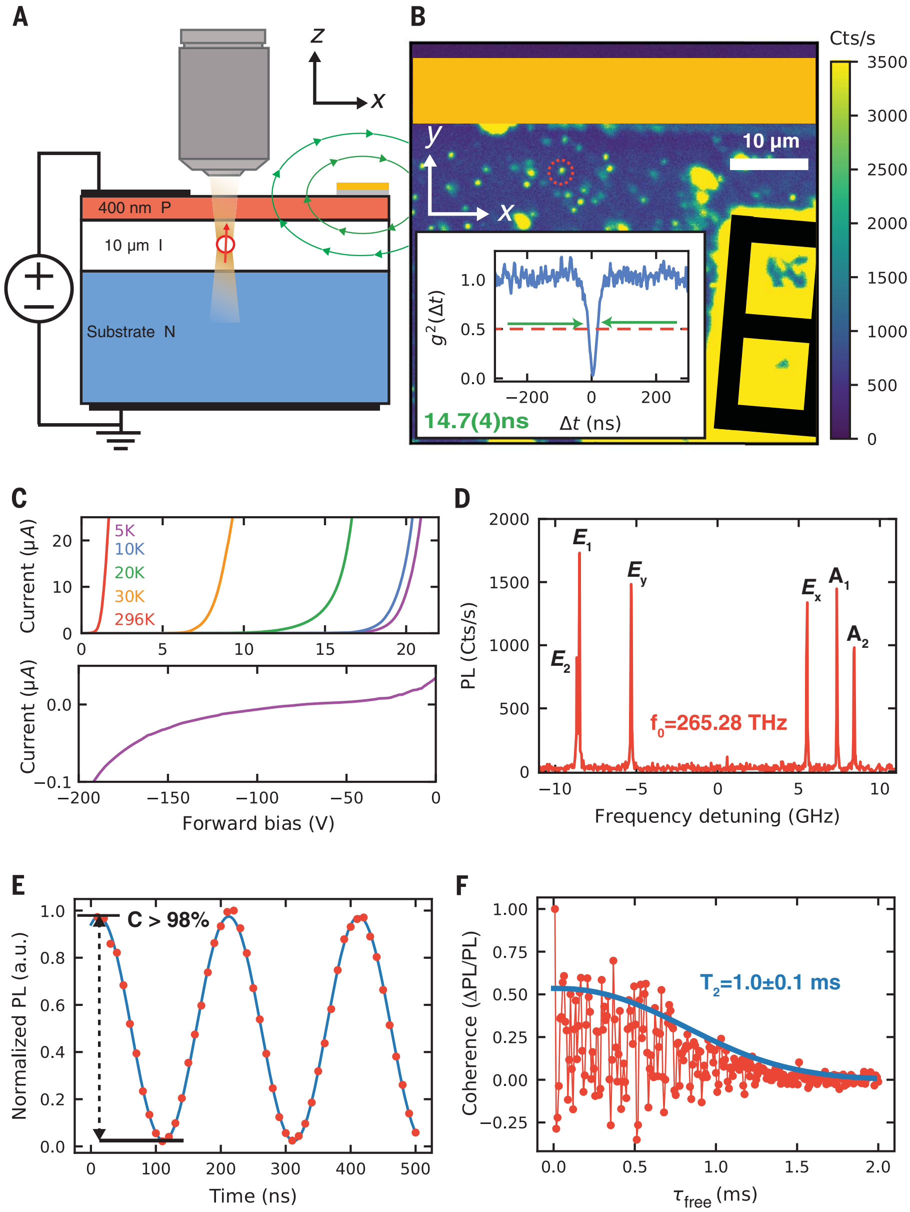


Fig. 1. Isolation of single VV^0 in a commercially grown semiconductor device. (**A**) Schematic of the device geometry. (**B**) Spatial PL scan of an example device showing isolated emitters (example circled in red) confirmed by autocorrelation (inset) showing $g^{(2)}$ (0) < 0.5 (red line). Extracted emitter lifetime is 14.7 ± 0.4 ns (green arrows). Gate and microwave stripline features are drawn and color coded as in (A). Cts, counts. (**C**) Top: Current–voltage (*I*–*V*) curves of the device at various temperatures; bottom: low-temperature reverse bias behavior. C, contrast. (**D**) PLE spectrum of a single (*kk*) divacancy at 270 V of reverse bias. (**E**) Optically detected Rabi oscillations of a single (*kk*) VV^0 with >98% contrast (fit in blue) using resonant initialization and readout. a.u., arbitrary units. (**F**) Hahn-echo decay of a single (*kk*) VV^0 in the diode. Rabi, Hahn, and $g^{(2)}$ data are taken at 270 V of reverse bias and at ~240 Gauss at T = 5 K.

electric field sensor, enabling field mapping in these working devices with sensitivities of ~100 $(V/m)/\sqrt{Hz}$ or better, which is competitive with state-of-the-art spin- and charge-based electrometry techniques (*32*, *43*–*46*).

Reducing spectral diffusion using charge depletion

Uncontrolled fluctuating electrical environments are a common problem in spin systems, where they can cause dephasing (*25*), as well as in quantum emitters, where they result in spectral diffusion of the optical structure and lead to large, inhomogeneous broadening. For example, adding and removing just a single electron charge 100 nm away causes shifts of

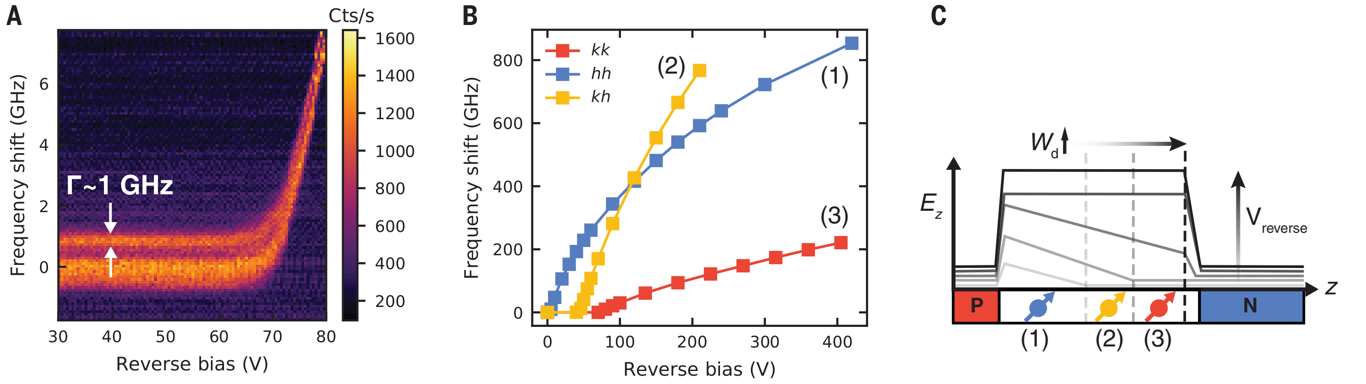


Fig. 2. Stark shifts in p-i-n diode. (**A**) Low-field Stark tuning of a single (*kk*) defect showing a turn-on behavior for the Stark shifts and a narrowing with voltage. This threshold is the same as that in Fig. 4A. These scans contain the lower branch (E_1, E_2, and E_y) where the linewidth of E_y is ~1 GHz and E_1 and E_2 are unresolved. The PLE lines show no shifting down to zero bias. (**B**) High-field Stark shifts of multiple example defects (located at various depths and positions in the junction) showing >100 GHz shifts. (**C**) Schematic electric field distribution and depletion region width (W_d) in the diode for increasing reverse bias. Location in the junction can determine the local field experienced by the defects in (B). The error bars in (B) are smaller than the point size. All data were obtained at T = 5 K.

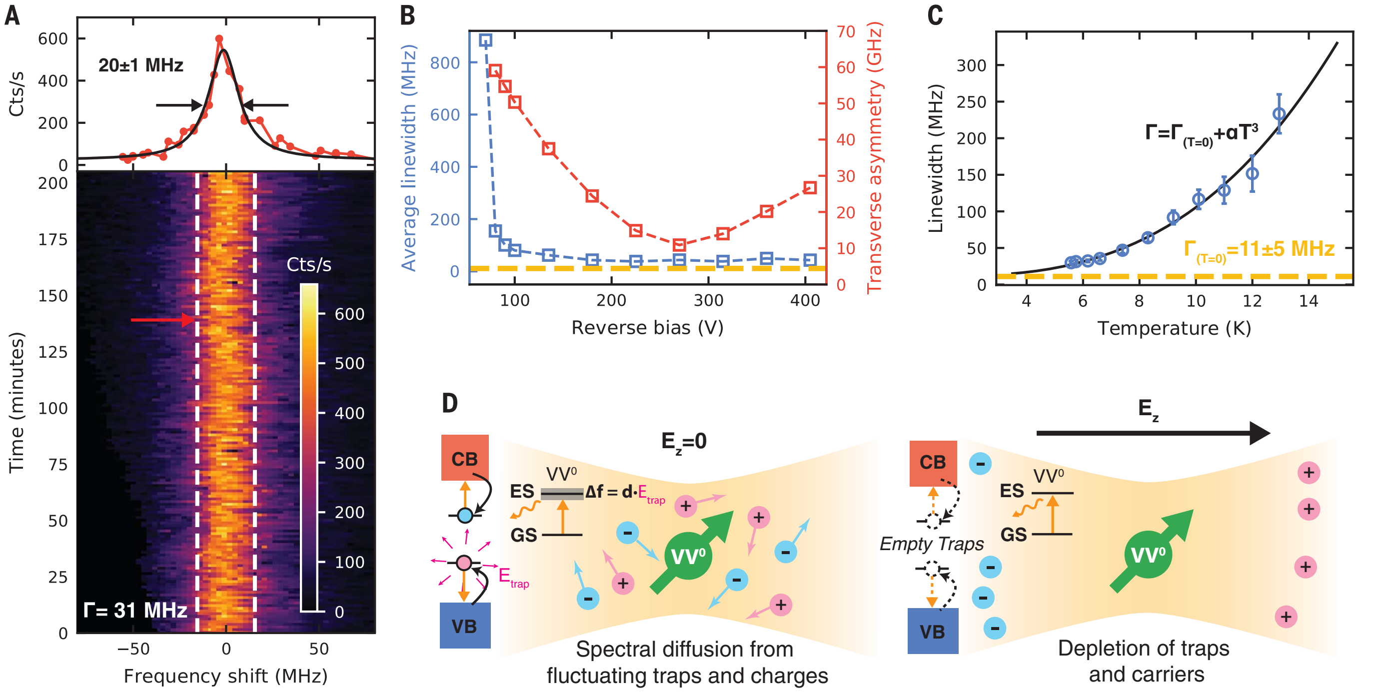


Fig. 3. Optical linewidth narrowing by tuning the electrical environment of a solid-state emitter. (**A**) Multiple PLE sweeps taken over 3.5 hours of the E_x line showing small residual spectral diffusion (fitted inhomogeneous linewidth of 31 ± 0.4 MHz). The red arrow corresponds to the single scan shown with a fitted linewidth of ~20 MHz. (**B**) Comparison of the average linewidth of all orbitals (blue) and defect transverse asymmetry (red) with respect to applied reverse bias. The yellow line is the lifetime limit. (**C**) Temperature dependence of the linewidth. A free power law fit gives an exponent of 3.2 ± 0.3. Constraining the fit to a T^3 relation, we extract a zero temperature linewidth of 11 ± 5 MHz (yellow line). Errors on the plot represent a 95% confidence interval. (**D**) Model for the effect of charge depletion on spectral diffusion in the illuminated volume (yellow). To the left of each diagram is a schematic band diagram with the relevant transitions. CB, conduction band; VB, valence band; GS, ground state; ES, excited state. Errors for the fits values in (A) and (C) represent 1 SD. All data are from a single (*kk*) VV^0. In (B), the laser power is slightly higher than in (A), causing some broadening. For (A) and (C), the E_x line is shown at 270 V of reverse bias. Data in (A) and (B) were obtained at T = 5 K.

~100 MHz for the optical fine structure of VV^0 (fig. S2). Previous work (*24*) has shown that by doing an exhaustive search through many defects in a specially grown material, one can find defects with lines as narrow as 80 MHz (typically 100 to 200 MHz or larger); however, this is still much larger than the Fourier lifetime limit of ~11 MHz (*24*). In bulk intrinsic commercial material, the narrowest linewidths are significantly broadened to 130 to 200 MHz or greater (*24*) (fig. S3). Overall, spectral diffusion has been a notoriously difficult outstanding challenge for nearly all quantum emitters in the solid state.

Here, we introduce a technique for mitigating spectral diffusion. We demonstrate that by applying electric fields in our device, we deplete the charge environment of our defect and obtain single-scan linewidths of 20 ± 1 MHz (Fig. 3A) without the need for an exhaustive search. This reduction in PLE linewidth has a different voltage dependence than the transverse asymmetry in the defect, thus eliminating reduced mixing as a possible mechanism for narrowing (Fig. 3B). The temperature dependence of the linewidth is roughly consistent with a T^3 scaling at these low temperatures (*47*) [fitted exponent 3.2 ± 0.3 and a zero-temperature linewidth of 11 ± 5 MHz (*32*)]. Although the dominant temperature scaling may change at lower temperature, this trend hints at a possible explanation for the remaining broadening and is consistent with a temperature-limited linewidth. Furthermore, the observed line is extremely stable, with a fitted inhomogeneous broadening of 31 ± 0.4 MHz averaged for >3 hours (Fig. 3A). This stability over time, narrowness, tunability, and photostability demonstrates the effectiveness of engineering the charge environment with doped semiconductor structures for creating ideal and indistinguishable quantum emitters.

At zero bias, the linewidth in our samples is much higher than in bulk material (~1 GHz; Fig. 2A). We attribute this to a greater presence of traps and free carriers (under illumination). Thus, in these samples, the observed narrowing corresponds to an improvement in the linewidth by a factor of >50. We speculate that a combination of this charge-depletion technique with lower sample temperatures, a lower-impurity material, and further annealing could enable measurement of consistent transform-limited linewidths (*13*, *48*). This use of charge depletion for creating spectrally narrow optical interfaces (Fig. 3D) could be widely applicable to other experiments in SiC or to other solid-state emitters such as quantum dots (*49*, *50*). Indeed, by applying the same techniques developed here to intrinsic SiC materials, lines as narrow as ~21 MHz have been observed (*40*). Crucially, these results demonstrate that depleting local charge environments can transform a very noisy electric environment into a clean one, turning materials containing unwanted impurities into ideal hosts for quantum emitters.

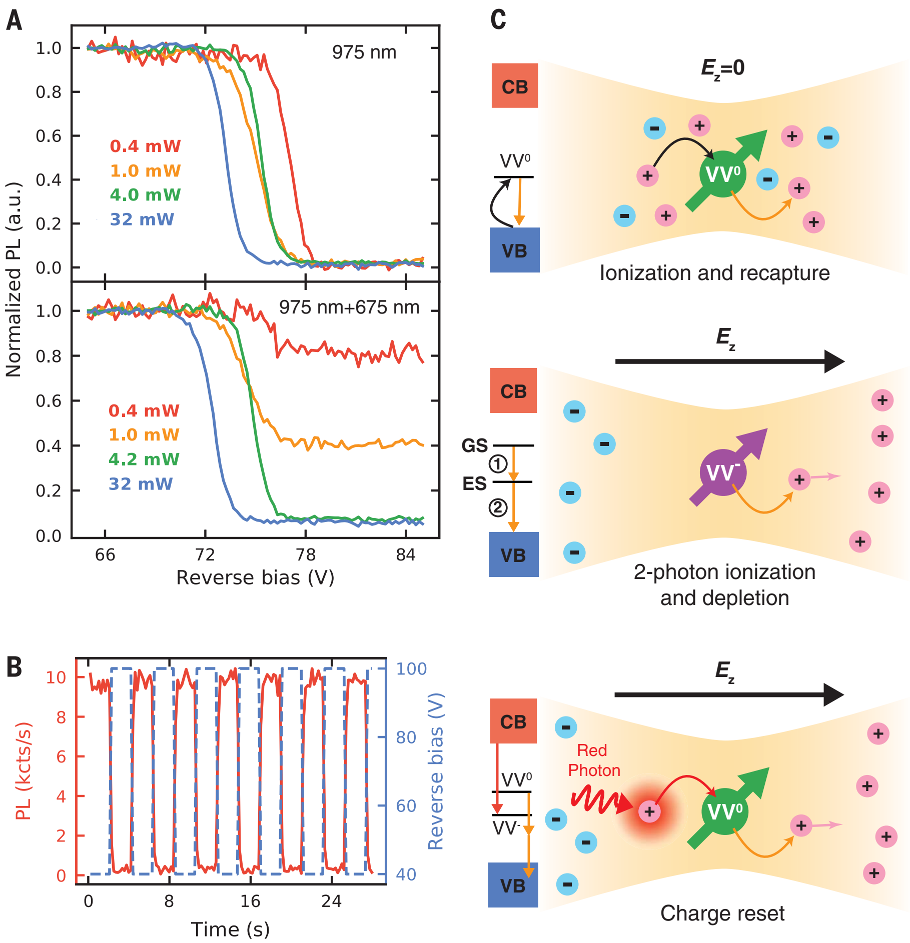


Fig. 4. Electrical and optical charge control of a single VV^0. (**A**) Voltage and power dependence of the photoluminescence of a single (*kk*) VV^0 with 975 nm excitation (top) and with additional 188 μW of 675 nm illumination (bottom), showing a sharp threshold under reverse bias. With high 975-nm power, the two-photon ionization process dominates and the PL signal is low. (**B**) By controlling the voltage in time (blue), the emission from the single (*kk*) defect is switched on and off (red). (**C**) Top: Model of rapid ionization and recapture at zero electric field (top). Middle: Two-photon ionization and formation of a depletion region under reverse bias. Bottom: Charge reset under applied electric field using red light (bottom). All data were obtained at T = 5 K. kcts, kilocounts.

Charge gating and photodynamics of single defects

Our observation of large Stark shifts and linewidth narrowing relies on understanding and controlling charge dynamics under electric fields. To achieve this, we studied the stability of the observed single defects under electrical bias. This allowed a careful investigation of the charge dynamics of single VV^0 under illumination, from which we developed an efficient charge-reset protocol. In our experiments, we observed that with 975 nm off-resonant light, the PL drops substantially once the reverse bias is increased past a threshold voltage (Fig. 4A). This threshold varies between defects, which is expected given differences in the local electric field stemming from variations in position, depth, and local charge-trap density. We attribute the PL reduction to photoionization to an optically "dark" charge state (*18*). We used this effect to create an electrically gated single-photon source (*51*–*53*) in which emission is modulated in time with a gate voltage (Fig. 4B) (*10*). The threshold voltage has a slight hysteresis (fig. S4) and laser power dependence (Fig. 4A), suggesting that trapped charges may play a role (*9*, *54*). The electric field dependence of the photoionization could also be used to extend sensitive electrometry techniques (*46*) to the single-defect regime, and controlled ionization of the spin can extend the coherence of nuclear registers (*27*). The threshold for Stark

shifts (Fig. 2A) corresponds approximately to the same voltage where significant photobleaching occurs when using off-resonant excitation. This links the sharp photoionization threshold in Fig. 4A to the presence of moderate electric fields and the onset of carrier depletion.

A possible explanation for this voltage-dependent PL is that at zero electric field, illumination constantly photoionizes the VV^0 and other nearby traps. However, the divacancy rapidly captures available free carriers, returning it to the neutral charge state. Under applied field, carrier drift depletes the illuminated region of charges. Thus, when a VV^0 photoionization event occurs in this depleted environment, no charges are available for fast recapture, resulting in a long-lived dark state (Fig. 4C).

Past studies have shown that PL is enhanced in ensembles by repumping the charge with higher-energy laser colors (*18*, *55*, *56*). We extended this work to the single-defect regime by applying various illumination energies and studying single-defect photodynamics at 90 V of reverse bias (past the threshold voltage of ~75 V of reverse bias for this defect). We observed under resonant illumination that PL quickly dropped to zero and did not recover, indicating that 1131 nm (1.09 eV) light [resonant with the ZPL of a (*kk*) VV^0] ionizes the defect, but does not reset the charge state. However, after applying higher-energy light (e.g., 688 nm), the charge was returned to a bright state even with <1 nW of applied power. This "repump" of the defect charge state is vital for restoring PL for ionized or charge unstable VV^0 in SiC (Fig. 4A) and is essential to observe the effects discussed in the previous sections (Fig. 4C).

When both near-infrared (NIR) resonant (1131 nm) and red (688 nm, 1.8 eV) light were applied to the defect, alternating between the bright (VV^0) and dark (VV^+ or VV^-) charge states resulted in a blinking behavior. From this blinking (fig. S5), we extract photoionization and repumping rates of the defect (*57*). We first examined the ionization rate of a single VV^0 (Fig. 5A) and observed that the power dependence was quadratic below defect saturation (exponent m = 2.05 ± 0.2) and linear at higher powers (m = 0.99 ± 0.07). Our observed data provide evidence for a two-photon process to VV^- (*32*) suggested in previous ensemble studies (*18*, *56*) and are less consistent with a recently proposed three-photon model converting to VV^+ (*35*, *55*). Thus, we conclude that the dark state caused by NIR resonant excitation is VV^-. Further study of the spin dependence of this ionization may lead to the demonstration of spin-to-charge conversion in VV^0.

Similarly, we studied the charge-reset kinetics by varying the power of the repumping laser. We found a near-linear power law with m = 0.98 ± 0.02 (Fig. 5B). This linear dependence of the repumping rate can be described by two potential models. One possibility is that the dark charge state is directly one-photon ionized by repump laser. The other possible explanation is that nearby traps are photoionized by this color and the freed charges are captured by the divacancy to convert back to the bright state. By varying the color of this reset laser, we found repumping to be most efficient at ~710 nm (1.75 eV), suggesting a particular trap-state energy or a possible defect absorption resonance (*58*, *59*) (Fig. 5C). Overall, we observed negligible ionization from the optimal red repump laser and no observable reset rate from the resonant laser. This results in fully deterministic optical control of the defect charge state [for discussion, see (*32*)], allowing for high-fidelity charge-state initialization for quantum-sensing and communications protocols.

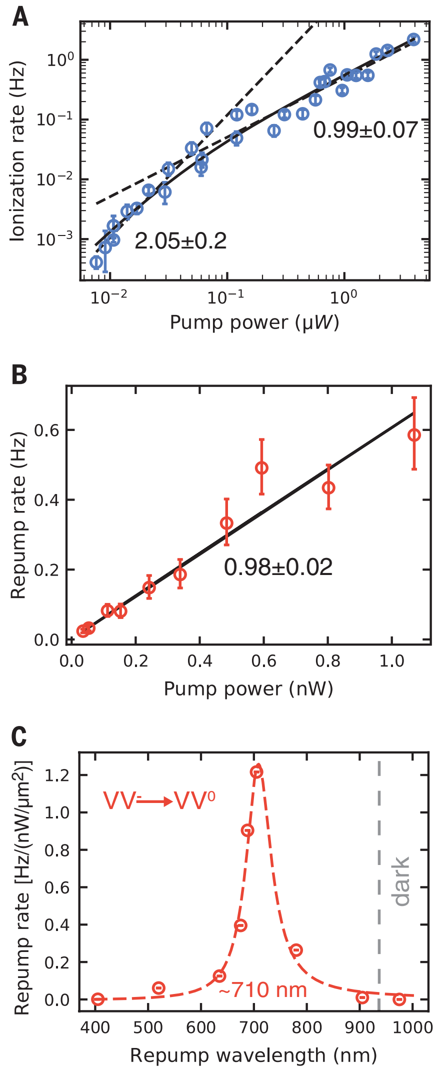


Fig. 5. Ionization and charge reset rates for VV^0. (**A**) Dependence of the ionization rate on resonant laser power. Low- and high-power regime fits (black dotted lines) and their power laws (m = 2.05 ± 0.2 and 0.99 ± 0.07, respectively) are shown. The solid black line shows a full model fit. (**B**) Repump power dependence of the 688 nm laser showing a linear exponent of m = 0.98 ± 0.02. Fluctuations in the polarization or power of the laser limit the true error. (A) and (B) were taken at 90 V of reverse bias. (**C**) Repumping rate as a function of illumination wavelength at 270 V of reverse bias with a Lorentzian fit centered around 710 nm. With wavelengths longer than 905 nm (and at these powers), no PL is observed and the defect is "dark." All error bars represent 95% confidence intervals from the fit of the raw data from a single (*kk*) VV^0. All data were obtained at T = 5 K.

Conclusions and outlook

The electrical tuning of the environment demonstrated here constitutes a general method that could be applicable to various quantum emitters in semiconductors in which spectral diffusion or charge stability is an issue (*60*) or electric field fluctuations limit spin coherence (*25*, *32*). Furthermore, using our p-i-n diode as a testbed to study charge dynamics, we have developed a technique to perform deterministic optical control of the charge state of single divacancies under electric fields (*61*).

The techniques presented here will be vital to achieving single-shot readout and entanglement in VV^0 by enabling charge control and enhancing photon indistinguishability, suggesting doped semiconductor structures as ideal quantum platforms for defects. This work also enables high-sensitivity measurement of nanoscale electric fields and charge distributions in working devices (*43*) and facilitates spin-to-charge conversion (*19*) for enhanced quantum-sensing and electrical readout protocols (*20*). Finally, the introduction of VV^0 into classical SiC semiconductor devices such as diodes, MOSFETs (metal-oxide-semiconductor field-effect transistors), and APDs (avalanche photodiodes), for example, may enable the next generation of quantum devices.

REFERENCES AND NOTES

1. C. Bonato *et al.*, *Nat. Nanotechnol.* **11**, 247–252 (2016).
2. G. Waldherr *et al.*, *Nature* **506**, 204–207 (2014).
3. D. D. Awschalom, R. Hanson, J. Wrachtrup, B. B. Zhou, *Nat. Photonics* **12**, 516–527 (2018).
4. L. Robledo *et al.*, *Nature* **477**, 574–578 (2011).
5. K. D. Jöns *et al.*, *Phys. Rev. B* **96**, 075430 (2017).
6. H. Bernien *et al.*, *Nature* **497**, 86–90 (2013).
7. D. Bluvstein, Z. Zhang, A. C. B. Jayich, *Phys. Rev. Lett.* **122**, 076101 (2019).
8. B. Kambs, C. Becher, *New J. Phys.* **20**, 115003 (2018).
9. L. C. Bassett, F. J. Heremans, C. G. Yale, B. B. Buckley, D. D. Awschalom, *Phys. Rev. Lett.* **107**, 266403 (2011).
10. C. F. de las Casas *et al.*, *Appl. Phys. Lett.* **111**, 262403 (2017).
11. W. Pfaff *et al.*, *Science* **345**, 532–535 (2014).
12. V. M. Acosta *et al.*, *Phys. Rev. Lett.* **108**, 206401 (2012).
13. Y. Chu *et al.*, *Nano Lett.* **14**, 1982–1986 (2014).
14. J. Wolters, N. Sadzak, A. W. Schell, T. Schröder, O. Benson, *Phys. Rev. Lett.* **110**, 027401 (2013).

15. T. Murai *et al*., *Appl. Phys. Lett.* **112**, 111903 (2018).
16. S. B. van Dam *et al*., *Phys. Rev. B* **99**, 161203 (2019).
17. K. Beha, A. Batalov, N. B. Manson, R. Bratschitsch, A. Leitenstorfer, *Phys. Rev. Lett.* **109**, 097404 (2012).
18. G. Wolfowicz *et al*., *Nat. Commun.* **8**, 1876 (2017).
19. B. J. Shields, Q. P. Unterreithmeier, N. P. de Leon, H. Park, M. D. Lukin, *Phys. Rev. Lett.* **114**, 136402 (2015).
20. P. Siyushev *et al*., *Science* **363**, 728–731 (2019).
21. W. F. Koehl, B. B. Buckley, F. J. Heremans, G. Calusine, D. D. Awschalom, *Nature* **479**, 84–87 (2011).
22. D. J. Christle *et al*., *Nat. Mater.* **14**, 160–163 (2015).
23. P. V. Klimov, A. L. Falk, D. J. Christle, V. V. Dobrovitski, D. D. Awschalom, *Sci. Adv.* **1**, e1501015 (2015).
24. D. J. Christle *et al*., *Phys. Rev. X* **7**, 021046 (2017).
25. M. Kim *et al*., *Phys. Rev. Lett.* **115**, 087602 (2015).
26. M. S. Grinolds *et al*., *Nat. Nanotechnol.* **9**, 279–284 (2014).
27. M. Pfender *et al*., *Nano Lett.* **17**, 5931–5937 (2017).
28. M. Widmann *et al*., *Appl. Phys. Lett.* **112**, 231103 (2018).
29. P. V. Klimov, A. L. Falk, B. B. Buckley, D. D. Awschalom, *Phys. Rev. Lett.* **112**, 087601 (2014).
30. M. W. Doherty *et al*., *Phys. Rev. X* **6**, 041035 (2016).
31. M. Niethammer *et al*., Coherent electrical readout of defect spins in 4H-SiC by photo-ionization at ambient conditions. arXiv:1903.12236 [cond-mat.mes-hall] (28 March 2019).
32. See the supplementary materials.
33. M. Widmann *et al*., *Nat. Mater.* **14**, 164–168 (2015).
34. N. T. Son *et al*., *Phys. Rev. Lett.* **96**, 055501 (2006).
35. A. Beste, D. E. Taylor, D. A. Golter, C. W. Lai, *Phys. Rev. B* **98**, 214107 (2018).
36. M. Shimizu *et al*., *Appl. Phys. Express* **11**, 033004 (2018).
37. K. Ohno *et al*., *Appl. Phys. Lett.* **101**, 082413 (2012).
38. J. R. Maze *et al*., *New J. Phys.* **13**, 025025 (2011).
39. S. J. Whiteley *et al*., *Nat. Phys.* **15**, 490–495 (2019).
40. K. C. Miao *et al*., *Sci. Adv.* **5**, eaay0527 (2019).
41. J. Forneris *et al*., *Phys. Rev. Appl.* **10**, 014024 (2018).
42. S. Wengerowsky, S. K. Joshi, F. Steinlechner, H. Hübel, R. Ursin, *Nature* **564**, 225–228 (2018).
43. D. A. Broadway *et al*., *Nat. Electron.* **1**, 502–507 (2018).
44. A. N. Vamivakas *et al*., *Phys. Rev. Lett.* **107**, 166802 (2011).
45. F. Dolde *et al*., *Nat. Phys.* **7**, 459–463 (2011).
46. G. Wolfowicz, S. J. Whiteley, D. D. Awschalom, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7879–7883 (2018).
47. N. R. Jungwirth *et al*., *Nano Lett.* **16**, 6052–6057 (2016).
48. S. O. Hruszkewycz *et al*., *Phys. Rev. Mater.* **2**, 086001 (2018).
49. H. Thyrrestrup *et al*., *Nano Lett.* **18**, 1801–1806 (2018).
50. M. C. Löbl *et al*., *Phys. Rev. B* **96**, 165440 (2017).
51. C. Chakraborty, L. Kinnischtzke, K. M. Goodfellow, R. Beams, A. N. Vamivakas, *Nat. Nanotechnol.* **10**, 507–511 (2015).
52. C. Schreyvogel, V. Polyakov, R. Wunderlich, J. Meijer, C. E. Nebel, *Sci. Rep.* **5**, 12160 (2015).
53. H. Kato *et al*., *Appl. Phys. Lett.* **102**, 151101 (2013).
54. F. J. Heremans, G. D. Fuchs, C. F. Wang, R. Hanson, D. D. Awschalom, *Appl. Phys. Lett.* **94**, 152102 (2009).
55. D. A. Golter, C. W. Lai, *Sci. Rep.* **7**, 13406 (2017).
56. B. Magnusson *et al*., *Phys. Rev. B* **98**, 195202 (2018).
57. N. Aslam, G. Waldherr, P. Neumann, F. Jelezko, J. Wrachtrup, *New J. Phys.* **15**, 013064 (2013).
58. M. Bockstedte, F. Schütz, T. Garratt, V. Ivády, A. Gali, *npj Quant. Mater.* **3**, 31 (2018).
59. P. Siyushev *et al*., *Phys. Rev. Lett.* **110**, 167402 (2013).
60. A. V. Kuhlmann *et al*., *Nat. Phys.* **9**, 570–575 (2013).
61. M. Widmann *et al*., *Nano Lett.* **19**, 7173–7180 (2019).
62. C. P. Anderson *et al*., Dataset for: Electrical and optical control of single spins integrated in scalable semiconductor devices, Zenodo (2019); https://doi.org/10.5281/zenodo.3523353.

ACKNOWLEDGMENTS

We thank E. O. Glen, S. Bayliss, D. J. Christle, P. V. Klimov, and P. J. Duda for helpful experimental suggestions and assistance and A. Gali, G. Galli, M. E. Flatté, D. R. Candido, and B. Magnusson for insightful discussions and theoretical understanding. Careful reading by F. J. Heremans supported manuscript preparation. We thank *Quantum Opus* for their assistance with SNSPDs. **Funding:** This work made use of the UChicago MRSEC (NSF DMR-1420709) and Pritzker Nanofabrication Facility, which receives support from the SHyNE, a node of the NSF's National Nanotechnology Coordinated Infrastructure (NSF ECCS-1542205). C.P.A., A.B., K.C.M., G.W., P.J.M., A.L.C., and D.D.A. were supported by AFOSR FA9550-14-1-0231 and FA9550-15-1-0029, DARPA D18AC00015KK1932, NSF EFRI EFMA-1641099, and ONR N00014-17-1-3026. C.P.A. was supported by the Department of Defense through the NDSEG Program. T.O. was supported by KAKENHI (17H01056 and 18H03770). J.U.H was supported by the Swedish Energy Agency (43611-1). N.T.S. was supported by the Swedish Research Council (VR 2016-04068) and the Carl Tryggers Stiftelse för Vetenskaplig Forskning (CTS 15:339). J.U.H. and N.T.S. were also supported by the Knut and Alice Wallenberg Foundation (KAW 2018.0071). **Author contributions:** C.P.A and A.B. conceived the experiments, fabricated the devices, performed the measurements, and analyzed the data. A.B. and K.C.M developed the experimental setup. A.L.C. assisted in device fabrication. H.A. and T.O. performed the electron irradiation. J.U.H. and N.T.S. assisted in growth and sample preparation of initial test devices used before the commercial samples. G.W. and P.J.M. measured initial devices. D.D.A. advised on all efforts. All authors contributed to the data analysis and manuscript preparation. **Competing interests:** A patent application has been filed relating to this work. P.J.M. is a paid consultant to ARCH Venture Partners. **Data and materials availability:** The data can be accessed at Zenodo (*62*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1225/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 and S2
References (*63–86*)

6 May 2019; accepted 5 November 2019
10.1126/science.aax9406

REPORT

SOLID-STATE PHYSICS

Direct determination of mode-projected electron-phonon coupling in the time domain

M. X. Na[1,2]*, A. K. Mills[1,2]*, F. Boschini[1,2], M. Michiardi[1,2,3], B. Nosarzewski[4], R. P. Day[1,2], E. Razzoli[1,2], A. Sheyerman[1,2], M. Schneider[1,2], G. Levy[1,2], S. Zhdanovich[1,2], T. P. Devereaux[4], A. F. Kemper[5], D. J. Jones[1,2]†, A. Damascelli[1,2]†

Ultrafast spectroscopies have become an important tool for elucidating the microscopic description and dynamical properties of quantum materials. In particular, by tracking the dynamics of nonthermal electrons, a material's dominant scattering processes can be revealed. Here, we present a method for extracting the electron-phonon coupling strength in the time domain, using time- and angle-resolved photoemission spectroscopy (TR-ARPES). This method is demonstrated in graphite, where we investigate the dynamics of photoinjected electrons at the $\overline{K}$ point, detecting quantized energy-loss processes that correspond to the emission of strongly coupled optical phonons. We show that the observed characteristic time scale for spectral weight transfer mediated by phonon-scattering processes allows for the direct quantitative extraction of electron-phonon matrix elements for specific modes.

The concept of the electronic quasiparticle as proposed by Landau (*1*) is essential to the modern understanding of condensed matter physics. Among the plethora of interactions relevant to solid-state systems, electron-phonon coupling (EPC)—which is related to phenomena ranging from resistivity in normal metals to conventional [Bardeen-Cooper-Schrieffer (BCS)] superconductivity and charge-ordered phases (*2*, *3*)—has been a persistent subject of interest. Although strong EPC is desirable in systems such as BCS superconductors (*4*, *5*), it is deleterious for conductivity in normal metals, curtailing the application of many compounds as room-temperature electronic devices (*6*).

Given the important role of the electron-phonon interaction in relation to both conventional and quantum materials, extensive theoretical and experimental efforts have been devoted to determining the strength and anisotropy of EPC. Although ab initio calculations are powerful, they rely on complex approximations that require precise experimental data to benchmark their validity (*7*). Inelastic scattering experiments—such as Raman spectroscopy (*8*), electron energy loss spectroscopy (*9*), inelastic x-ray (*10*), and neutron scattering (*11*)—are able to access EPC for specific phonon modes yet are integrated over all electronic states. Angle-resolved photoemission spectroscopy (ARPES), in contrast, can access the strength of EPC via phonon-mediated renormalization effects for specific momentum-resolved electronic states, as revealed by kinks in the electronic band dispersion (*12*–*16*). However, extraction of EPC strength from these kinks requires accurate modeling of the bare band dispersion and of the electronic self-energy, which can prove to be a formidable challenge either because of insufficient sensitivity and experimental resolution (*17*), or because of too-strong and/or compounded many-body interactions (*18*, *19*). In addition, the interpretation of spectroscopic features is often complicated by the fact that they may be attributed to several different many-body interactions (*20*–*22*).

Alternative and possibly more powerful approaches might come from the extension of ARPES into the time domain [time-resolved ARPES (TR-ARPES)], which has already provided deep insights into the relaxation channels of hot electronic distributions, in which EPC plays a major role (*23*–*28*). TR-ARPES performed with 6-eV sources has enabled detailed study of low-energy many-body

[1]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. [2]Quantum Matter Institute, Vancouver, BC V6T 1Z4, Canada. [3]Max Planck Institute for Chemical Physics of Solids, 01187 Dresden, Germany. [4]Department of Materials Science and Engineering, Stanford Institute for Materials and Energy Sciences, Stanford, CA 94305, USA. [5]Department of Physics, North Carolina State University, Raleigh, NC 27695, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: djjones@physics.ubc.ca (D.J.J.); damascelli@physics.ubc.ca (A.D.)

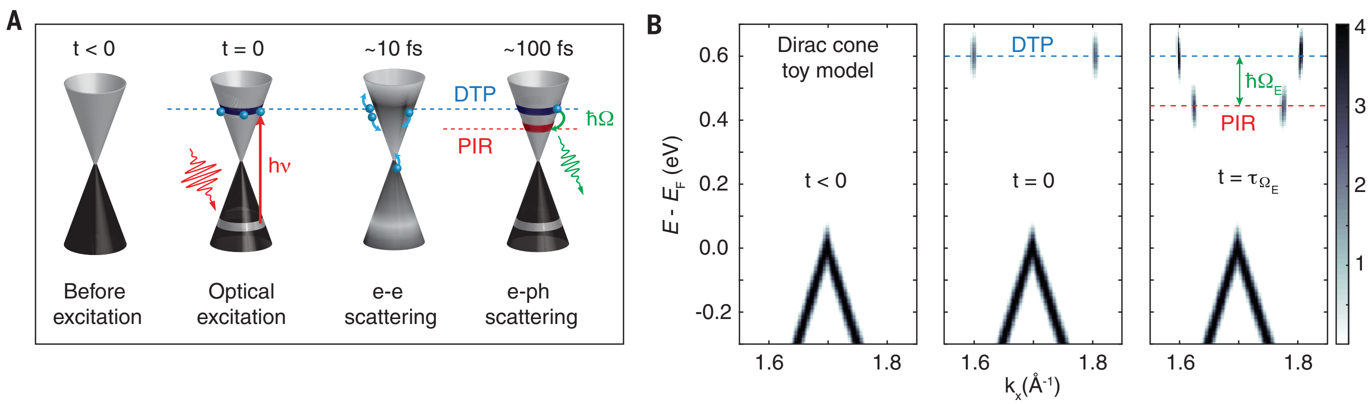


Fig. 1. Toy model of optical injection and scattering processes on the Dirac cone. (**A**) Sketch of the Dirac cone and electron dynamics. Black (gray) indicates occupied (unoccupied) states. During optical excitation, electrons from the lower cone are promoted to the upper cone through a vertical transition (red arrow), creating a direct-transition peak (DTP). Electrons subsequently relax and scatter through electron-electron (e-e) and electron-phonon (e-ph) processes on time scales of 10 and 100 fs, respectively. The former (e-e) broadens the DTP, whereas the latter creates a phonon-induced replica (PIR) by the emission of a phonon. (**B**) Simulation of the transient TR-ARPES intensity for a Dirac cone pumped with 1.2 eV, including a retarded e-ph interaction with a phonon of energy $\hbar\Omega_E$ (*39*). At time t = 0, the DTP feature is observed at $E_{\Omega_E} = 0.6$ eV; at $t = \tau_{\Omega_E}$, the PIR is observed at $E_{DTP} - \hbar\Omega_E$. The intensity of DTP and PIR features is enhanced (×8) for visualization purposes.

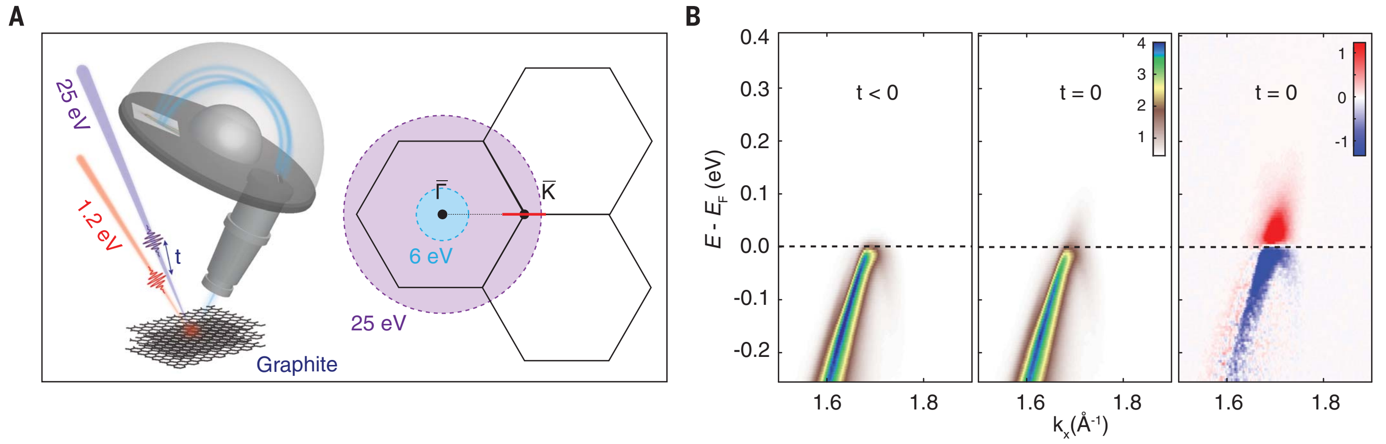


Fig. 2. Electron dynamics measured by TR-ARPES in graphite. (**A**) The experimental setup, along with the 2D-projected Brillouin zone of graphite. Blue (purple) circle indicates the range of momenta accessible to 6-eV (25-eV) photons. We measure along the $\overline{\Gamma} - \overline{\mathrm{K}}$ direction, cut shown in red. (**B**) TR-ARPES measurements acquired with a 25-eV probe and 1.19-eV pump. Sample temperature is 50 K before pump arrival. The unpumped dispersion (t < 0) is shown in the left panel. The pumped ARPES map at zero delay and the differential map are shown in the middle and right panels, respectively. Owing to fast thermalization processes, the signal of DTP and PIR cannot be observed by simple visual inspection on a linear colormap.

phenomena; although this has provided a fresh perspective on superconducting gap dynamics of cuprate superconductors (*29–31*), electron-phonon interaction in bulk FeSe (*32*), and surface-state dynamics in topological materials (*26*, *33*), low photon energies have limited these studies to a small region of the Brillouin zone (BZ). ARPES systems using high-harmonic sources have extended the accessible momenta beyond the first BZ, but heretofore they have focused on the high-energy–scale electron dynamics on the order of 10 fs (*25*, *27*, *34*, *35*), as energy resolutions have yet to reach the standards achieved by 6-eV systems.

Here, we explore a paradigm for the TR-ARPES study of transient spectral features at large momenta, made possible by a femtosecond high-harmonic source designed with specific emphasis on energy resolution (*36*). The experimental strategy is as follows: We begin by injecting electrons into specific unoccupied states by optical excitation. As the hot electrons relax, we track specifically the transfer of spectral weight from these photoexcited states to lower-energy states via emission of a phonon with energy $\hbar\Omega_{\mathbf{q},\nu}$, where $\mathbf{q}$ and ν denote the phonon momentum and branch, respectively. The time constant extracted for this transfer of spectral weight ($\tau_{\mathbf{q},\nu}$) can then be directly related to the electron-phonon contribution to the self-energy for the phonon involved as

$$\frac{1}{\tau_{\mathbf{q},\nu}} = \frac{2\pi}{\hbar}\langle g^2_{\mathbf{q},\nu}\rangle D(E - \hbar\Omega_{\mathbf{q},\nu}) \quad (1)$$

where $\hbar$ is the reduced Planck constant, E is the energy of the direct optical transition, $\langle g^2_{\mathbf{q},\nu}\rangle$ is the square of the mode-projected electron-phonon matrix element averaged over the states populated by optical excitation, and $D(E)$ is the electronic density of states (DOS) (*7*, *37*, *38*) [for derivation, see (*39*)]. Below, we show that this allows us to measure $\langle g^2_{\mathbf{q},\nu}\rangle$, gaining insight on the strength of the scattering process as well as the energy and momenta involved.

To track the transfer of spectral weight, the initial (photoinjected) states and final states must be unambiguously defined and located. This is easiest on a small Fermi surface, such as that of graphene, where phase-space scattering restrictions limit the number of initial and final states. To visualize the aforementioned electron-phonon scattering process, we simulate the pump-probe experiment using a Dirac dispersion as a toy model and calculate the transient ARPES spectra in response to optical excitation for the case of a single strongly coupled Einstein phonon mode of energy $\hbar\Omega_{\mathbf{q},\nu} = \hbar\Omega_{\mathrm{E}}$. This model (Fig. 1B) uses a two-time Green's function formalism on the Keldysh contour for a multiorbital system. At t = 0, the system is excited with a 1.2-eV pulse, which promotes electrons via direct optical excitation to 0.6 eV, observed experimentally as the direct-transition peak (DTP). Then, on the characteristic time scale of $\tau_{\mathbf{q},\nu} = \tau_{\Omega_{\mathrm{E}}}$, the electron-phonon interaction leads to relaxation of the photoinjected electron population via the emission of phonons, resulting in the creation of a phonon-induced replica (PIR) at an energy $E_{\mathrm{PIR}} = 0.6$ eV$-\hbar\Omega_{\mathrm{E}}$. In contrast to the toy model, other scattering processes (such as electron-electron scattering) compete with this electron-phonon scattering process in the actual experiment, resulting in a smaller PIR and a larger background. In comparing Fig. 1B and Fig. 2B, we do not see the predicted DTP and PIR features in the intensity map. However, we show that these features are plainly visible in the momentum-integrated energy distribution curves (Fig. 3).

In this study, we perform the experiment on graphite, which has the same ideal phase-space restrictions as its monolayer counterpart but does not require consideration of substrate coupling, which in graphene is known to affect both electronic and phononic structure, as well as EPC (*40–42*). The low-energy electronic structure of single-crystal graphite consists of two gapless, nearly two-dimensional (2D) Dirac-like bands at the BZ corners (similar to graphene) as well as a second set of bands that disperse along the c axis ($k_\perp$ in our experimental geometry, see Fig. 4A) (*43–45*). In addition, graphite electrons are well known to couple to optical phonons at $\overline{\Gamma}$ and $\overline{\mathrm{K}}$ (*10*, *46–50*) and have been extensively studied both theoretically and experimentally (*8*, *24*, *28*), making graphite an ideal benchmark system for the application of this time-resolved technique.

Previous time-resolved experiments have shown that the time (energy) scale of the electron-phonon scattering process is on the order of 100 fs (100 meV). Therefore, observation of this transient spectral signature in TR-ARPES demands a balance of time and energy resolution. Achieving the system resolution requirements at photon energies needed to reach the $\overline{\mathrm{K}}$ point (>20 eV, assuming a maximum detection angle of 60°; see Fig. 2A for the BZ range covered by photons of different energy) was made

viable by the development of a new cavity-based high-harmonic source (*36*). We select the 21st harmonic (25 eV) from the high-harmonic spectrum for photoemission, with an overall time (energy) resolution of 190 fs (22 meV) and a repetition rate of 60 MHz. The unpumped ARPES map of the Dirac-like dispersion along $\overline{\Gamma}-\overline{K}$ is shown in the left panel of Fig. 2B, where only one branch of the cone is observed as a consequence of photoemission matrix elements (*51*, *52*). In the middle and right panels of Fig. 2B, we show, respectively, the pumped ARPES spectra at zero delay and the differential map (obtained by subtracting the equilibrium map from the map at zero pump-probe delay). The data were measured under perturbative excitation by a 1.19-eV pump pulse with an incident fluence of 18 $\mu J/cm^2$, where red (blue) color indicates a transient increase (decrease) of photoemission intensity. The time scales of the anticipated primary scattering processes after optical excitation are sketched in Fig. 1A. After the creation of the DTP above the Fermi level (E_F), electrons decay into a thermal distribution, mainly via electron-electron (e-e) and electron-phonon (e-ph) scattering events. Because the e-e scattering processes are about an order of magnitude faster than e-ph scattering (*53*), we should observe a rapid buildup of photoemission intensity at the Fermi energy. This can render the observation of the DTP and PIR nontrivial, requiring a careful analysis of features above the hot-electron background.

In Fig. 3A, we display the momentum-integrated energy distribution curve ($\int$EDC $d\mathbf{k}$) along the $\overline{\Gamma}-\overline{K}$ direction (open circles). We stress that the $\int$EDC $d\mathbf{k}$ is proportional to the occupied DOS along the selected momentum cut shown in Fig. 2A (*39*). Filled circles in Fig. 3A represent the $\int$EDC $d\mathbf{k}$ after removal of a biexponential background given by the thermal electronic distribution (near E_F) and nonthermal e-e scattering processes (near 0.6 eV) (*39*). Once this background is removed, the $\int$EDC $d\mathbf{k}$ directly exposes the transient peaks, which can be fitted with five Lorentzians of the same width (Fig. 3A). We can immediately identify the prominent peak at 0.6 eV as DTP_1, which was anticipated in the toy model (Fig. 1B) and is associated with the optical transition from the π_2-to-π_3 band in Fig. 4A. The other peaks—as we show in more detail below—are a combination of PIRs and other DTPs, which arise from the second set of electronic bands (π_1, π_4) that disperse in $k_\perp$. We confirm these transitions in Fig. 4A with a calculation of the optical joint DOS for graphite (*54*), adapted from a tight-binding model in (*45*), for a pump photon energy of 1.19 eV. The possible transitions along the $\overline{\Gamma}-\overline{K}$ cut are shown in Fig. 4A. Whereas the π_2-to-π_4 transition is outside the range of our data, the three lower DTPs fall exactly in the energy range we expect. The resulting momentum-integrated optical joint DOS is shown in Fig. 4B, along with the energy position of the five fitted peaks from Fig. 3A.

To illustrate the DTP-to-PIR scattering process, we focus on the time evolution of the three most prominent peaks, shown in Fig. 3B; for a discussion of the DTP_2-PIR_2 pair, see (*39*). The combined time and energy resolution of our source allows for a detailed study of the transient evolution of the DTP and the PIR, given by the amplitude of the Lorentzians in Fig. 3C. Despite being only 50 meV apart, the dynamics of the light-blue and red peaks are markedly different. The population of the light-blue peak is only slightly delayed with respect to the dark-blue DTP_1 and is identified with a direct optical excitation (π_1-to-π_4 band in Fig. 4A, labeled DTP_2), with the temporal delay being a consequence of energy-dependent electron lifetime (*55*, *56*). In contrast, the population of the red peak is delayed by $\Delta t = 47 \pm 9$ fs, too large to be compatible with optical excitation. This population instead corresponds to the simulated PIR in Fig. 1B, where the energy of the phonon involved is $\hbar\Omega_{\mathbf{q},\nu} = E_{DTP} - E_{PIR} = 0.165 \pm 0.011$ eV.

The solid curves in Fig. 3C are the result of a phenomenological rate-equation model describing the transfer of spectral weight between the DTP and the PIR pairs (*39*). The population of electrons in the (dark and light) blue DTP are governed by rate equations involving three terms: population by a 120-fs pump pulse, energy-dependent thermalization of the excited state population to the hot-electron bath (τ_{th}), and energy-dependent phonon-mediated decay of the excited state

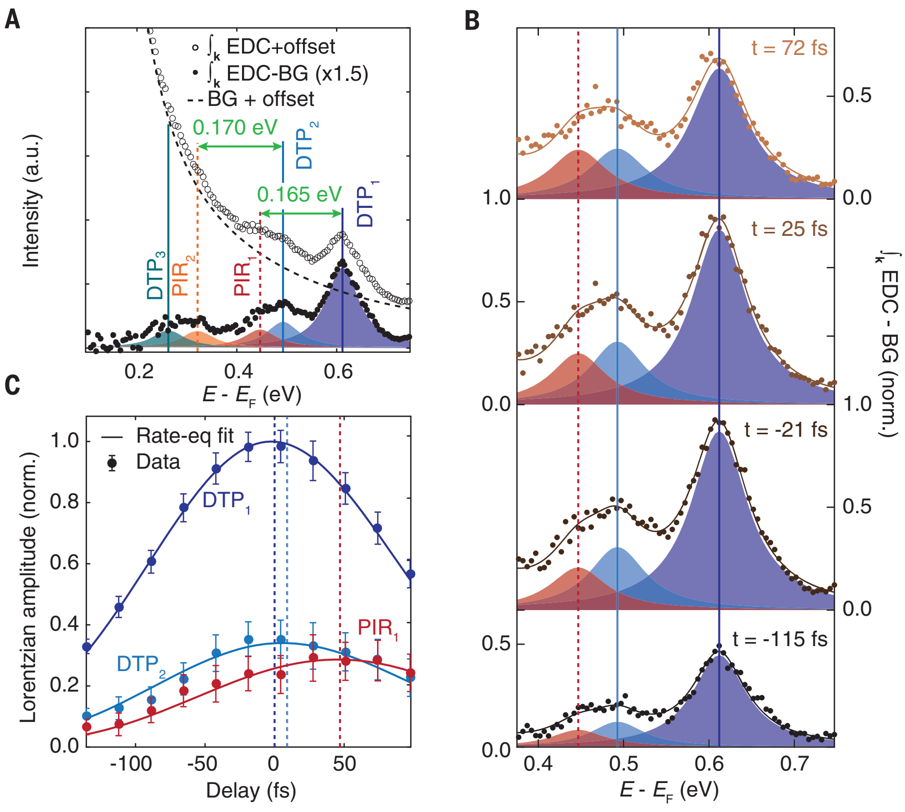


Fig. 3. Time dependence of photoinduced excitations in graphite. Sample temperature is 50 K before pump arrival. (**A**) Open circles display the momentum-integrated energy distribution curve $\int$EDC $d\mathbf{k}$, where signal is integrated in momentum along the $\overline{\Gamma}-\overline{K}$ direction. The subtraction of the biexponential hot-electron background (BG) highlights a series of peaks (filled circles), which are a combination of DTPs and PIRs. Solid (dashed) lines indicate the fitted position of DTP (PIR) peaks. The phonon energies extracted between the DTP_1-PIR_1 and DTP_2-PIR_2 pairs are 0.165 eV and 0.170 eV, respectively, as indicated by the green arrows. (**B**) Evolution of the most prominent peaks. Dark (light) blue corresponds to DTP_1 (DTP_2), red corresponds to PIR_1. The amplitudes are indicative of the population of electrons in each state. (**C**) The Lorentzian amplitude for each peak shown in (B) is plotted as a function of time. Dashed lines indicate the peak delay: DTP_2 (PIR_1) is delayed 9 fs (47 fs) with respect to DTP_1. Solid curves indicate the electronic occupation for the specified states derived from the rate-equation model fit. The transfer of spectral weight from DTP_1 to PIR_1 is associated to an e-ph scattering time constant $\tau_{\mathbf{q},\nu} = 174 \pm 35$ fs.

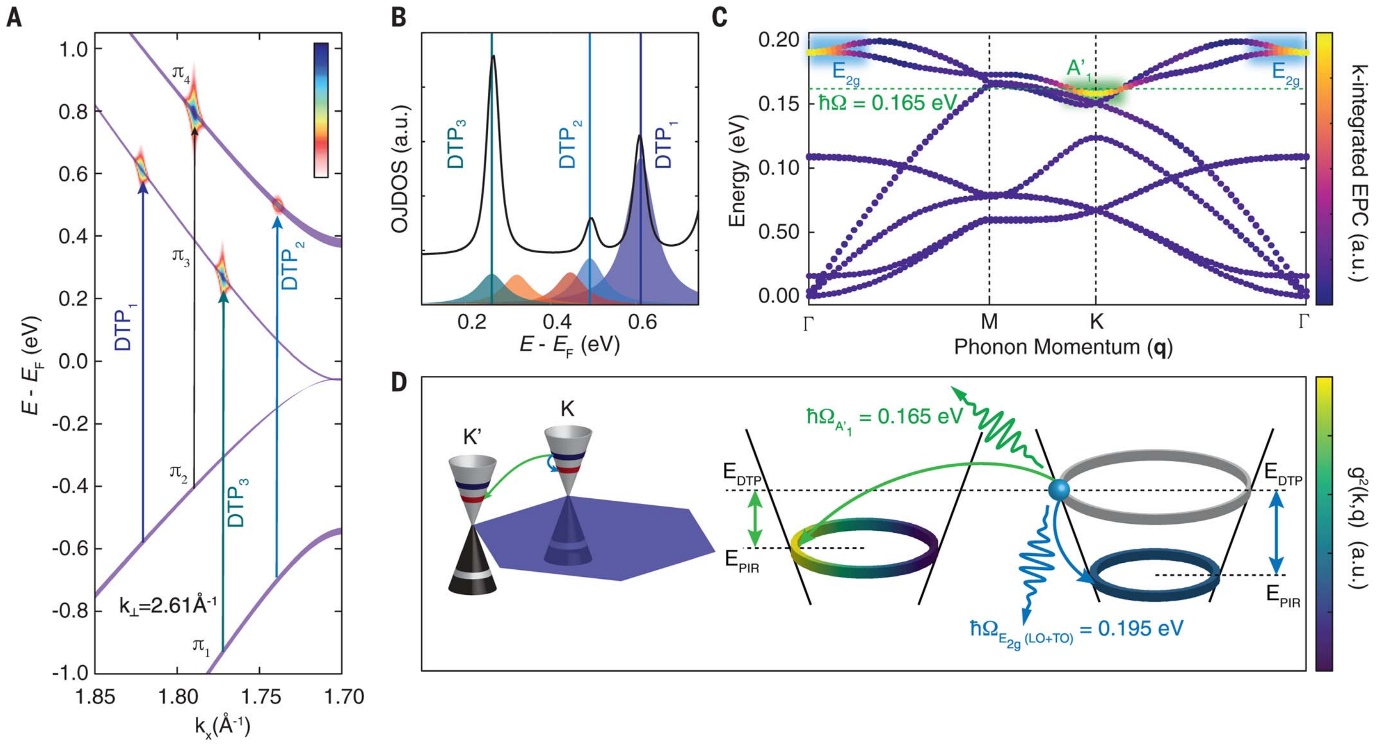


Fig. 4. Calculation of the optical joint DOS and e-ph coupling. (**A**) Momentum-resolved optical joint DOS, extracted using a modified tight-binding model from (*39*, *45*, *54*), showing the available optical transitions between the bands of graphite for a pump photon energy of 1.19 eV, integrated around $k_\perp = 2.61 Å^{-1}$. This value is fitted using an inner potential of V_0 = 16.4 ± 0.1 eV (*44*). (**B**) Integrated optical joint DOS (OJDOS) along the $\overline{\Gamma} - \overline{K}$ direction. The peaks extracted from Fig. 3A are also shown, with the location of the fitted DTP overlaid in solid lines, displaying good agreement with the optical joint DOS. (**C**) The phonon-dispersion of graphite as calculated from DFT (*38*). Colors represent the strength of the total EPC (i.e., integrated over all electronic momenta) (*39*). The dominant mode A'_1 (E_{2g}) is highlighted in green (blue) (*46*). (**D**) Calculation of the $\langle g^2_{\mathbf{k},\mathbf{q}} \rangle$ (*39*). Case one: electron scattering with an A'_1 mode (green) with momentum ~K from a specific state (indicated by the blue sphere) in the 0.61-eV energy contour at K to states in the 0.45-eV contour at K′. Case two: electron scattering with an E_{2g} (blue) mode with momentum ~0 from a specific state in the 0.61-eV energy contour at K to states in the 0.41-eV contour at K. $\langle g^2_{A'_1} \rangle$ and $\langle g^2_{E_{2g}} \rangle$ are obtained by integrating over both the constant energy contour at E_{PIR} and the electron position along the contour at E_{DTP}. Colors indicate that the value of $\langle g^2_{A'_1} \rangle$ is twice that of the LO and TO combined $\langle g^2_{E_{2g}} \rangle$ (see text for precise values). The scattering process from K′ to K (not shown) is identical.

population ($\tau_{\mathbf{q},\nu}$). This latter term transfers spectral weight from the DTP$_i$ to the PIR$_i$, which lose electrons to the same thermalization and phonon-mediated decay terms. The resultant temporal evolution is then convolved with a Gaussian with a full width at half maximum of 150 fs to account for the pulse duration of the photoemission probe. With this simple model, we find that a thermalization constant of τ_{th} = 56 ± 16 fs and an e-ph decay constant of $\tau_{\mathbf{q},\nu}$ = 174 ± 35 fs reproduce well the delay and relative population of the nonthermal signatures. In addition, the extracted $\tau_{\mathbf{q},\nu}$ is consistent with ab initio calculations and estimates for e-ph scattering time in previous time-resolved studies (*25*, *27*, *28*, *53*, *57*). Because we are coupling to a single-phonon mode of energy $\hbar\Omega_{\mathbf{q},\nu} = 0.165 \pm 0.011$ eV, we can directly relate this time constant to the mode-projected e-ph matrix element via Eq. 1, using an electronic DOS $[D(E - \hbar\Omega_{\mathbf{q},\nu}) = 0.0241\ \mathrm{eV}^{-1}]$ calculated from the tight-binding model in Fig. 4A. From this, we obtain a value of $\langle g^2_{\mathbf{q},\nu} \rangle = 0.050 \pm 0.011\ \mathrm{eV}^2$.

We now assign the observed PIR to scattering by a specific phonon mode by comparing the extracted $\hbar\Omega_{\mathbf{q},\nu}$ against the phonon dispersion of graphite calculated by density functional theory (DFT) in Fig. 4C. The colors represent the EPC integrated over all electronic momenta and indicate strong coupling for the E_{2g} mode at Γ and the A'_1 mode at K. The latter is the phonon mode associated with the DTP-PIR pair we observe, as its energy matches the 0.165 eV we extract (green dashed line). Given that it has momentum K, this mode is associated with the intervalley scattering of electrons between states at $\overline{K}$ and $\overline{K'}$. We consider this scattering process explicitly for a single electron in Fig. 4D (green arrow). Starting from an initial state i on the constant energy contour E_{DTP_1}, we calculate the matrix element $g^2_{\mathbf{k},\mathbf{q}}$ for scattering events leading to the final state f on the constant energy contour $E_{\mathrm{DTP}_1} - \hbar\Omega_{A'_1}$, such that $\mathbf{k}_f - \mathbf{k}_i = \mathbf{q}$ is fulfilled. This mode-projected calculation gives a value of $\langle g^2_{A'_1} \rangle = 0.040\ \mathrm{eV}^2$, in agreement with the experimental value of $\langle g^2_{\mathbf{q},\nu} \rangle = \langle g^2_{A'_1} \rangle = 0.050 \pm 0.011\ \mathrm{eV}^2$ that we previously extracted from the rate-equation fits to the experimental data. In addition to the A'_1 mode, Fig. 4C suggests that the Γ – E_{2g} modes [longitudinal (LO) and transverse (TO) modes] are also expected to be strongly coupled; however, considering the scattering process as before, we extract for the degenerate LO and TO modes a total coupling $\langle g^2_{E_{2g}} \rangle = 0.023\ \mathrm{eV}^2$, which corresponds to a time constant of >300 fs. This coupling is ~50% that of the A'_1 mode, which is consistent with previous theoretical considerations (*58*). Thus, the PIR associated with emission of the

$\Gamma - E_{2g}$ phonons would not be visible above the hot-electron background in our experiment.

In this work, we have shown a proof-of-principle extraction of the mode-projected e-ph matrix element $\langle g^2_{A'_1} \rangle$ in graphite. We remark that $\langle g^2 \rangle$ is a fundamental quantity, defined as the momentum average of the change in the electronic Hamiltonian in response to the ionic displacements of a phonon (eq. S13). In particular, $\langle g^2 \rangle$ is independent of doping and specified for a well-defined set of initial and final electronic states and a well-defined bosonic mode (in this case, the A'_1 optical phonon with momentum K in graphite). Nonetheless, it is instructive to estimate the EPC constant $\lambda = 2\langle g^2 \rangle D(E_F)/(\hbar\Omega)$ for comparison with other approaches (*7*). Commonly seen in relation to the critical temperature in superconductivity, the quantity λ is doping-dependent and integrated over all bosonic and electronic degrees of freedom. Therefore, caution must be applied in comparing the two quantities. In pristine and low-doping graphene or graphite systems, the vanishing DOS at the Dirac point of graphene (crossing point of graphite) makes extraction of λ very difficult, with reported values ranging from 4×10^{-4} to 1.1 (*20*, *47*, *59*–*63*), while DFT predicts $\lambda < 0.05$ (*17*). Even forgoing doping dependence, the large range of reported λ values clearly illustrates the difficulty ARPES has in extracting the EPC of this particular system. In this work, we extract $\langle g^2_{A'_1} \rangle$ for the DTP state at 0.6 eV, corresponding to a value of the mode-projected EPC $\lambda_{A'_1} = 0.0182 \pm 0.004$. This value characterizes a system where the Fermi level is doped up to 0.6 eV above the crossing point (the value at zero doping would be $\lambda_{A'_1} = 0.006 \pm 0.001$). We would ideally compare this value with that extracted by kink analysis in graphite, but the strong curvature of the bare-band dispersion (*61*) and the lack of studies at comparable doping make this particularly challenging. Thus, we instead compare this value to what is reported in (*20*), a doping-dependent study of λ in graphene that is additionally supported by DFT calculations (*17*). When the system is doped such that the Dirac point is 0.6 eV below E_F (corresponding to a carrier density of $n \approx 4 \times 10^{13}$ cm^{-2}), then $\lambda \approx 0.035$ is extracted. From this, we see that $\lambda_{A'_1} \approx \lambda/2$, which is consistent with the fact that $\lambda_{A'_1}$ captures only one of two strongly coupled modes in the system (the other being E_{2g}), whereas λ as extracted from kink analysis is integrated over all modes. Together, these results suggest that time-domain measurements have the ability to access the EPC in a precise, sensitive, and mode-projected way.

In principle, this technique is applicable to materials in which electrons are sufficiently strongly coupled to one or a few bosonic modes, such that distinct boson-induced replicas can be observed. Beyond graphite and graphene, quasi-2D materials such as transition-metal dichalcogenides feature gapped bands at the K and K′ points, with further restrictions of the phase space for scattering stemming from valley degrees of freedom, and would be ideal candidates for similar TR-ARPES studies. In addition, conventional and unconventional superconductors, such as MgB_2 or cuprate- or Fe-based superconductors, respectively, famously feature strong coupling to bosonic modes, which may drive electronic renormalizations (kinks). By monitoring quantized decay processes across the full BZ, this nonequilibrium technique will offer a distinct approach for studying the microscopic origin and momentum dependence of electron-boson coupling and its role in the emergence of superconductivity.

These results also prove that a fresh analytical perspective can be achieved in TR-ARPES by taking advantage of femtosecond sources that combine high photon energy (to access large electronic momenta) with high-energy resolution (to resolve the low-energy quasiparticle dynamics). With the development of tunable pumps, polarization control for pump and probe, and bandwidth control to balance the trade-off between energy and time resolution, a growing versatility will be available for TR-ARPES experiments. By monitoring the population of electrons injected into momentum- and energy-selected states by direct optical excitation, it would be possible to formulate a series of studies on empty state dispersion, lifetime, decoherence, and electron-boson interactions in a wide range of quantum materials.

REFERENCES AND NOTES

1. L. D. Landau, *J. Exp. Theor. Phys.* **3**, 920 (1957).
2. M. Le Tacon *et al.*, *Nat. Phys.* **10**, 52–58 (2014).
3. T. P. Devereaux *et al.*, *Phys. Rev. X* **6**, 041019 (2016).
4. M. Strongin *et al.*, *Phys. Rev. Lett.* **21**, 1320–1323 (1968).
5. W. L. McMillan, *Phys. Rev.* **167**, 331–344 (1968).
6. M. Scheuch *et al.*, *Appl. Phys. Lett.* **99**, 211908 (2011).
7. F. Marsiglio, J. P. Carbotte, in *Superconductivity: Conventional and Unconventional Superconductors*, K. H. Bennemann, J. B. Ketterson, Eds. (Springer-Verlag, 2008), pp. 73–162.
8. A. C. Ferrari, *Solid State Commun.* **143**, 47–57 (2007).
9. S.-i. Tanaka, K. Mukai, J. Yoshinobu, *Phys. Rev. B* **95**, 165408 (2017).
10. M. Mohr *et al.*, *Phys. Rev. B* **76**, 035439 (2007).
11. T. Yildirim *et al.*, *Phys. Rev. Lett.* **87**, 037001 (2001).
12. T. Valla, A. V. Fedorov, P. D. Johnson, S. L. Hulbert, *Phys. Rev. Lett.* **83**, 2085–2088 (1999).
13. S. LaShell, E. Jensen, T. Balasubramanian, *Phys. Rev. B* **61**, 2371–2374 (2000).
14. A. Lanzara *et al.*, *Nature* **412**, 510–514 (2001).
15. A. Damascelli, *Phys. Scr.* **2004**, 61 (2004).
16. J. Shi *et al.*, *Phys. Rev. Lett.* **92**, 186401 (2004).
17. M. Calandra, F. Mauri, *Phys. Rev. B* **76**, 205411 (2007).
18. N. J. Ingle *et al.*, *Phys. Rev. B* **72**, 205114 (2005).
19. C. N. Veenstra, G. L. Goodvin, M. Berciu, A. Damascelli, *Phys. Rev. B* **84**, 085126 (2011).
20. D. A. Siegel, C. Hwang, A. V. Fedorov, A. Lanzara, *New J. Phys.* **14**, 095006 (2012).
21. C. Zhang *et al.*, *Nat. Commun.* **8**, 14468 (2017).
22. F. Li, G. A. Sawatzky, *Phys. Rev. Lett.* **120**, 237001 (2018).
23. L. Perfetti *et al.*, *New J. Phys.* **10**, 053019 (2008).
24. Y. Ishida *et al.*, *Sci. Rep.* **1**, 64 (2011).
25. I. Gierz *et al.*, *Nat. Mater.* **12**, 1119–1124 (2013).
26. J. A. Sobota *et al.*, *J. Electron Spectrosc. Relat. Phenomena* **195**, 249–257 (2014).
27. A. Stange *et al.*, *Phys. Rev. B* **92**, 184303 (2015).
28. J. A. Yang, S. Parham, D. Dessau, D. Reznik, *Sci. Rep.* **7**, 40876 (2017).
29. L. Perfetti *et al.*, *Phys. Rev. Lett.* **99**, 197001 (2007).
30. C. L. Smallwood *et al.*, *Science* **336**, 1137–1139 (2012).
31. F. Boschini *et al.*, *Nat. Mater.* **17**, 416–420 (2018).
32. S. Gerber *et al.*, *Science* **357**, 71–75 (2017).
33. J. A. Sobota *et al.*, *Phys. Rev. Lett.* **108**, 117403 (2012).
34. T. Rohwer *et al.*, *Nature* **471**, 490–493 (2011).
35. F. Cilento *et al.*, *Sci. Adv.* **4**, eaar1998 (2018).
36. A. K. Mills *et al.*, *Rev. Sci. Instrum.* **90**, 083001 (2019).
37. M. Sentef *et al.*, *Phys. Rev. X* **3**, 041033 (2013).
38. T. Sohier, thesis, De L'Université Pierre et Marie Curie Spécialité, Paris, France (2016).
39. Materials and methods are available as supplementary materials.
40. S. Y. Zhou *et al.*, *Nat. Mater.* **6**, 770–775 (2007).
41. Y. Y. Wang *et al.*, *J. Phys. Chem. C* **112**, 10637–10640 (2008).
42. A. Allard, L. Wirtz, *Nano Lett.* **10**, 4335–4340 (2010).
43. J. C. Slonczewski, P. R. Weiss, *Phys. Rev.* **109**, 272–279 (1958).
44. A. Grüneis *et al.*, *Phys. Rev. Lett.* **100**, 037601 (2008).
45. C. M. Cheng *et al.*, *Appl. Surf. Sci.* **354**, 229–234 (2015).
46. S. Piscanec, M. Lazzeri, F. Mauri, A. C. Ferrari, J. Robertson, *Phys. Rev. Lett.* **93**, 185503 (2004).
47. S. Y. Zhou, G. H. Gweon, A. Lanzara, *Ann. Phys.* **321**, 1730–1746 (2006).
48. M. Lazzeri, C. Attaccalite, L. Wirtz, F. Mauri, *Phys. Rev. B* **78**, 081406 (2008).
49. R. P. Chatelain, V. R. Morrison, B. L. M. Klarenaar, B. J. Siwick, *Phys. Rev. Lett.* **113**, 235502 (2014).
50. M. J. Stern *et al.*, *Phys. Rev. B* **97**, 165416 (2018).
51. I. Gierz, J. Henk, H. Höchst, C. R. Ast, K. Kern, *Phys. Rev. B* **83**, 121408(R) (2011).
52. Y. Liu, G. Bian, T. Miller, T. C. Chiang, *Phys. Rev. Lett.* **107**, 166803 (2011).
53. S. Ulstrup *et al.*, *J. Phys. Condens. Matter* **27**, 164206 (2015).
54. R. P. Day, B. Zwartsenberg, I. S. Elfimov, A. Damascelli, *NPJ Quant. Mat.* **4**, 54 (2019).
55. S. Xu *et al.*, *Phys. Rev. Lett.* **76**, 483–486 (1996).
56. P. Narang, L. Zhao, S. Claybrook, R. Sundararaman, *Adv. Opt. Mater.* **5**, 1600914 (2017).
57. J. C. Johannsen *et al.*, *Phys. Rev. Lett.* **111**, 027403 (2013).
58. S. Butscher, F. Milde, M. Hirtschulz, E. Malić, A. Knorr, *Appl. Phys. Lett.* **91**, 203103 (2007).
59. K. Sugawara, T. Sato, S. Souma, T. Takahashi, H. Suematsu, *Phys. Rev. Lett.* **98**, 036801 (2007).
60. A. Bostwick, T. Ohta, T. Seyller, K. Horn, E. Rotenberg, *Nat. Phys.* **3**, 36–40 (2007).
61. C. S. Leem *et al.*, *Phys. Rev. Lett.* **100**, 016802 (2008).
62. S. Ulstrup *et al.*, *Phys. Rev. B* **86**, 161402(R) (2012).
63. F. Joucken *et al.*, *Phys. Rev. B* **93**, 241101(R) (2016).
64. M. Na, Replication Data for: Direct determination of mode-projected electron-phonon coupling in the time-domain, Version 1, Harvard Dataverse (2019); https://doi.org/10.7910/DVN/TFKS8Z.
65. M. Na, mengxing-n/graphite: v1.0.0, Version 1, Zenodo (2019); https://doi.org/10.5281/zenodo.3525476.

ACKNOWLEDGMENTS

We gratefully acknowledge M. Berciu, G. A. Sawatzky, A. Nocera, Z. Ye, and D. Manske for critical reading of the manuscript and useful discussions, as well as C. Gutierréz for guidance on figure presentation. **Funding:** This research was undertaken thanks in part to funding from the Max Planck-UBC-UTokyo Centre for Quantum Materials and the Canada First Research Excellence Fund, Quantum Materials and Future Technologies Program. This project is funded in part by the Gordon and Betty Moore Foundation's EPiQS Initiative, Grant GBMF4779 to A.D. and D.J.J.; the Killam, Alfred P. Sloan, and Natural Sciences and Engineering Research Council of Canada's (NSERC's) Steacie Memorial Fellowships (A.D.); the Alexander von Humboldt Fellowship (A.D.); the Canada Research Chairs Program (A.D.); NSERC, Canada Foundation for Innovation (CFI); British Columbia Knowledge Development Fund (BCKDF); and the CIFAR Quantum Materials Program. E.R. acknowledges support from the Swiss National Science Foundation (SNSF) grant P300P2_164649. A.F.K. acknowledges support by the National Science Foundation under grant DMR-1752713. B.N. and T.P.D. acknowledge funding from the Department of Energy, Basic Energy Sciences, Division of Materials Science. **Author contributions:** This study was conceived of by M.X.N., A.K.M., F.B., M.M., D.J.J., and A.D.; and

the experiment was carried out by M.X.N., A.K.M., F.B., and M.M., with experimental support from E.R., A.S., M.S., G.L., and S.Z., as well as theoretical contributions from R.P.D., B.N., T.P.D., and A.F.K. The enhancement cavity–based HHG laser source system was conceived of by D.J.J. and A.D. and was developed by A.K.M. and D.J.J. All authors worked together to interpret and model the results and to write the paper. D.J.J. and A.D. were responsible for overall project direction, planning, and management. **Competing interests:** The authors have no competing interests. **Data and materials availability:** Data shown in the main text and supplementary materials are available at the Harvard Dataverse (*64*). All experimental data and calculations shown in the main text and in the supplementary materials are accessible at Zenodo (*65*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1231/suppl/DC1
Materials and Methods
Analysis and Models
Theoretical Calculations
Figs. S1 to S7
Tables S1 to S3
References (*66–77*)
Movie S1

23 November 2018; accepted 5 November 2019
10.1126/science.aaw1662

CONSERVATION ECOLOGY

Extinction filters mediate the global effects of habitat fragmentation on animals

Matthew G. Betts[1]*†, Christopher Wolf[1]*†, Marion Pfeifer[2], Cristina Banks-Leite[3], Víctor Arroyo-Rodríguez[4], Danilo Bandini Ribeiro[5], Jos Barlow[6,7], Felix Eigenbrod[8], Deborah Faria[9], Robert J. Fletcher Jr.[10], Adam S. Hadley[1], Joseph E. Hawes[11], Robert D. Holt[12], Brian Klingbeil[13], Urs Kormann[1,14,15], Luc Lens[16], Taal Levi[1], Guido F. Medina-Rangel[17], Stephanie L. Melles[18], Dirk Mezger[19], José Carlos Morante-Filho[9,20], C. David L. Orme[3], Carlos A. Peres[21], Benjamin T. Phalan[22], Anna Pidgeon[23], Hugh Possingham[24,25], William J. Ripple[1], Eleanor M. Slade[26], Eduardo Somarriba[27], Joseph A. Tobias[3], Jason M. Tylianakis[28], J. Nicolás Urbina-Cardona[29], Jonathon J. Valente[1,30], James I. Watling[31], Konstans Wells[32], Oliver R. Wearn[33], Eric Wood[34], Richard Young[35], Robert M. Ewers[3]

Habitat loss is the primary driver of biodiversity decline worldwide, but the effects of fragmentation (the spatial arrangement of remaining habitat) are debated. We tested the hypothesis that forest fragmentation sensitivity—affected by avoidance of habitat edges—should be driven by historical exposure to, and therefore species' evolutionary responses to disturbance. Using a database containing 73 datasets collected worldwide (encompassing 4489 animal species), we found that the proportion of fragmentation-sensitive species was nearly three times as high in regions with low rates of historical disturbance compared with regions with high rates of disturbance (i.e., fires, glaciation, hurricanes, and deforestation). These disturbances coincide with a latitudinal gradient in which sensitivity increases sixfold at low versus high latitudes. We conclude that conservation efforts to limit edges created by fragmentation will be most important in the world's tropical forests.

Global biodiversity loss is occurring at more than 100 times the prehuman background extinction rate (*1*), and there is general consensus among scientists that most species' declines can be attributed to habitat loss (*2*, *3*). Nevertheless, the degree to which habitat fragmentation, defined as the spatial arrangement of remaining habitat, influences biodiversity loss has been a source of contention for over 40 years (*4*–*7*). Resolving this debate is important to conservation planning, which can entail designing the configuration of landscapes as well as spatially prioritizing areas for conservation (*8*). Forest fragmentation is particularly pressing given that 70% of Earth's remaining forest is within 1 km of the forest edge (*9*) and that fragmentation of the world's most intact forest landscapes—the tropics—is predicted to accelerate over the coming five decades (*10*).

The variation across taxa and regions in species' responses to fragmentation and edge effects in particular is central to the fragmentation debate (*6*, *11*, *12*). It is well known that life history and other ecological traits mediate species' responses to habitat edges (*13*), but the degree to which there are predictable geographical patterns in species' sensitivity has yet to be quantified across multiple taxa on a global scale.

Species' evolutionary histories can shape their capacity to respond to novel stressors. The extinction filter hypothesis predicts that species that have evolved and survived in high-disturbance environments should be more likely to persist in the face of new disturbances, including those of habitat loss and fragmentation (*14*). Further, more frequent disturbances could act as a barrier to sensitive species, preventing them from colonizing disturbance-prone regions. Disturbances often create edges, and in environments with frequent and large-scale disturbances, persistent species are more likely to be adapted to ubiquitous edge habitats. The extinction filter hypothesis is at least several decades old and has been suggested to apply in forest (*15*, *16*) and grassland systems (*14*). Both natural disturbances (such as wildfires and glaciation) and anthropogenic ones (such as logging, burning, and hunting) are thought to exert such evolutionary pressures (*14*). Nevertheless, there has been no global test of whether historical disturbance regimes can explain fragmentation effects.

We used 73 datasets collected worldwide containing 4489 species from four major taxa [2682 arthropods, 1260 birds, 282 herptiles (reptiles and amphibians), and 265 mammals] (Fig. 1, fig. S1, and tables S1 and S2) to provide a global test of the extinction filter hypothesis in forest ecosystems (*17*). In the presence of an extinction filter, species inhabiting a filtered landscape with high levels of disturbances over historical (evolutionary) time scales should be resilient to new disturbances—either because sensitive species have been driven locally extinct or because extant species have adapted to disturbance. Either mechanism would lead to a reduced prevalence of fragmentation-sensitive species in regions of the globe where disturbance has been historically common.

We used a recently developed approach to quantify the landscape-scale impacts of forest edges on biodiversity (*13*, *18*). By definition, habitat fragmentation for a given habitat amount leads to more, smaller patches, with a greater proportion of edge. We focus on landscape-scale variation in edge effects rather than the number of patches, because edge effects have long been known to have widespread effects on

[1]Forest Biodiversity Research Network, Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA. [2]School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK. [3]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK. [4]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México (UNAM), Campus Morelia, Antigua Carretera Patzcuaro no. 8701, Ex-Hacienda de San José de la Huerta, 58190 Morelia, Michoacán, Mexico. [5]Instituo de Biociências, Universidade Federal de Mato Grosso do Sul, Caixa Postal 549, 79070-900 Campo Grande, Brazil. [6]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. [7]Setor Ecologia, Departamento de Biologia, Universidade Federal de Lavras, 37200-000, Lavras, MG, Brazil. [8]Geography and Environmental Sciences, University of Southampton, Southampton SO17 1BJ, UK. [9]Applied Conservation Ecology Lab, Programa de Pós-graduação em Ecologia e Conservação, da Biodiversidade, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Salobrinho, 45662-000 Ilhéus, Bahia, Brazil. [10]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32611, USA. [11]Applied Ecology Research Group, School of Life Sciences, Anglia Ruskin University, Cambridge CB1 1PT, UK. [12]Department of Biology, University of Florida, Gainesville, FL 32611, USA. [13]School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA. [14]Swiss Ornithological Institute, Sempach, Switzerland. [15]Division of Forest Sciences, School of Agricultural, Forest and Food Sciences HAFL, Bern University of Applied Sciences, Zollikofen, Switzerland. [16]Ghent University, Department of Biology, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium. [17]Groupo de Biodiversidad y Conservación, Reptiles, Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Ciudad Universitaria, Edificio 425, Bogotá, Distrito Capital, Colombia. [18]Department of Chemistry and Biology, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada. [19]Department of Science and Education, Field Museum of Natural History, Chicago, IL 60605, USA. [20]Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Avenida Transnordestina, s/n - Novo Horizonte, 44036-900 Feira de Santana, Bahia, Brazil. [21]Centre for Ecology, Evolution and Conservation, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [22]Instituto de Biologia, Universidade Federal da Bahia, Salvador, 40170-115 Bahia, Brazil. [23]Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, 1630 Linden Drive, Madison, WI 53706, USA. [24]School of Biological Sciences, University of Queensland, St Lucia, Queensland, Australia. [25]The Nature Conservancy, Arlington, VA 22203, USA. [26]Asian School of the Environment, Nanyang Technological University, 62 Nanyang Dr., 637459 Singapore. [27]Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica. [28]School of Biological Sciences, University of Canterbury, Private bag 4800, Christchurch 8140, New Zealand. [29]Department of Ecology and Territory, School of Rural and Environmental Studies, Pontificia Universidad Javeriana, Bogota, Colombia. [30]Smithsonian Conservation Biology Institute, Migratory Bird Center, National Zoological Park, Washington, DC 20013, USA. [31]Department of Biology, John Carroll University, University Heights, OH 44118, USA. [32]Department of Biosciences, Swansea University, Swansea SA2 8PP, Wales, UK. [33]Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK. [34]Department of Biological Sciences, California State University Los Angeles, 5151 State University Drive, Los Angeles, CA 90032, USA. [35]Durrell Wildlife Conservation Trust, Les Augres Manor, Trinity, Jersey JE3 5BP, UK.
***Corresponding author. Email: matt.betts@oregonstate.edu (M.G.B); wolfch@oregonstate.edu (C.W)** †These authors contributed equally to this work.

biodiversity (*13*), and the approach we use comprehensively captures the process of fragmentation at the landscape scale (*17*) (figs. S2 and S3).

These data and methodology have been documented extensively elsewhere (*13*), so we present a brief overview relevant to our analysis. Each dataset contains a set of sample points within a fragmented forest region where abundances of one or more species from major taxonomic groups were sampled. We quantified two key aspects of edge effects: edge influence across the region and edge sensitivity of species. We quantified edge influence (EI) surrounding sample points on the basis of variation in percentage of forest cover (*13*, *17*). This metric accounts for the cumulative effects of multiple edges (including edge shape and patch size) that magnify the realized impact of edges on species. Edge sensitivity (*S*) is a population-specific measure of fragmentation sensitivity that ranges from 0.0 (no edge response) to 1.0 (high edge avoidance or preference). Because *S* does not distinguish between forest and matrix species or between edge avoidance and edge preference, we also used abundance, percentage of tree cover within 30 m of sample points, and EI to classify species as forest, nonforest matrix, or generalist habitat users and as core, edge, or no preference (*17*) (fig. S4). We did this by simulating sets of example abundances in each category (e.g., forest core) and then using a naïve Bayes classifier to estimate the most likely category for each actual species on the basis of abundance versus point cover and EI relationships. By definition, forest core species are those that are restricted to forest areas distant from the edge; hence, these species are sensitive to fragmentation of large patches into smaller ones (figs. S2 and S3). We used this classification as the basis for our statistical models, focusing on both the probability of forest species being classified as core and the probability of species being classified as forest, matrix, or generalist. For each study site, we assembled available data on forest fire severity (*19*), whether or not its location was glaciated in the last glacial maximum (*20*), whether or not it experienced tropical storms (*21*), and if historical anthropogenic forest loss at the site exceeded 50% (*3*, *17*) (Fig. 1).

Across all species combined, we found strong support for the extinction filter hypothesis explaining geographically variable sensitivity to forest edge. The odds of forest species being classified as forest core were 79.0% (95% confidence interval: 65.9 to 87.0%) lower in study regions that have experienced historically severe disturbances ($P < 0.001$) (Fig. 2 and table S3). A substantial 51.3% of forest species tended to avoid edges in low-disturbance regions, whereas only 18.1% of forest species in high-disturbance regions avoided edges (Fig. 2).

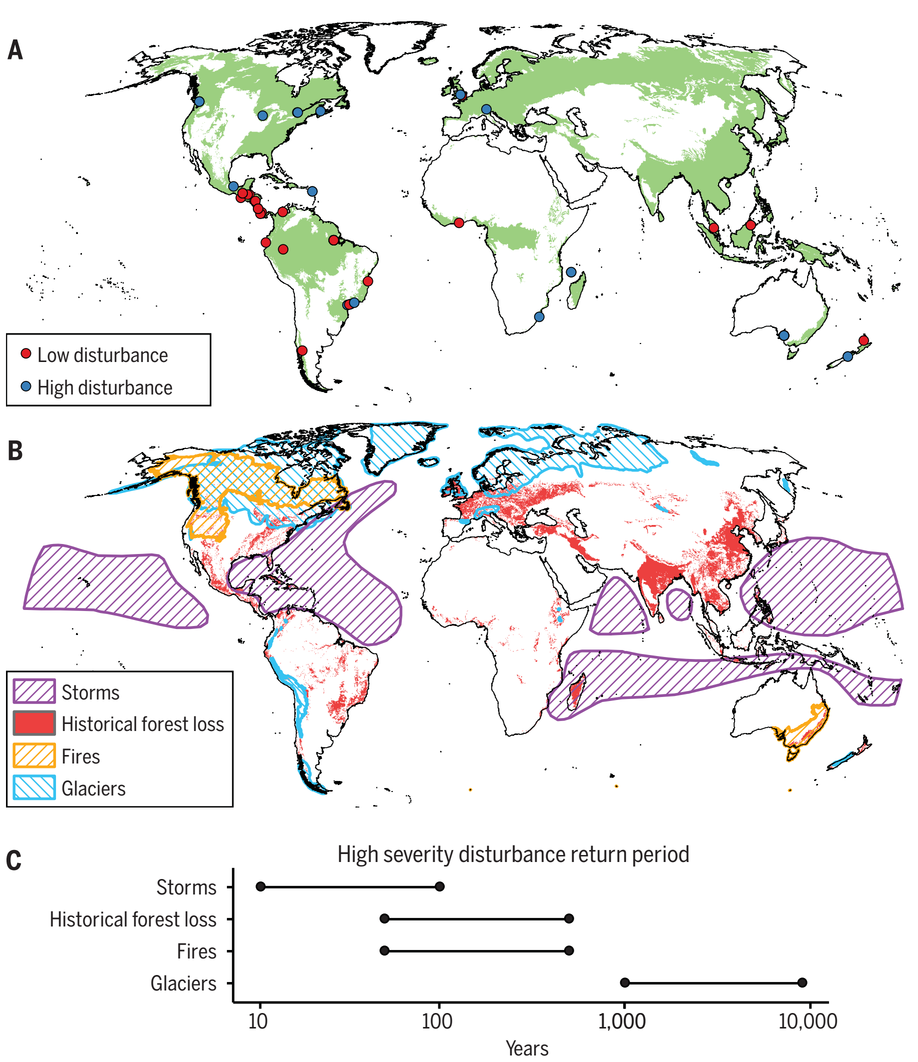


Fig. 1. Geographic distributions of sample study regions and historical disturbances. (**A**) Locations of the 35 BIOFRAG regions where the 73 datasets included in our analysis were collected. Areas that can support forests are shown in green. The regions are colored according to disturbance severity. (**B**) Distributions of historical disturbances: tropical storms, historical (long-term) deforestation, high-intensity crown fires, and glaciation. (**C**) Typical periods over which high-severity disturbances return to the same location.

Edge-sensitive species are therefore largely absent from communities in historically disturbed locations, suggesting that they have either disappeared from these regions or adapted to become less edge sensitive. This result was particularly strong for arthropods and birds, and the results were in the same direction for herptiles and mammals, though nonsignificant, likely owing to lower sample sizes. Results were stronger still when we considered the proportion of forest species as a function of disturbance severity. The odds of a species being forest associated versus being associated with other habitats were 729% (95% credible interval: 608 to 891%) higher in low-disturbance versus high-disturbance regions (fig. S5 and table S4).

Edge sensitivity (*S*) of forest core species tended to be 1.16 times as high in low-disturbance regions [$S = 0.660 \pm 0.004$ (standard error)] as in high-disturbance regions ($S = 0.568 \pm 0.004$). This effect size is considerable; species with values of $S > 0.75$ are found only within the forest interior far away from edges, whereas forest species with $S = 0.5$ are abundant up to the edges (*13*). Additionally, historical anthropogenic forest loss alone was substantially less effective at predicting the proportion of core species than either the combination of historical disturbances or natural disturbance alone (table S3). Thus, evolutionary responses and patterns of extinction of forest species in high-disturbance regions are not driven solely by anthropogenic habitat loss and fragmentation.

The effects of disturbance on edge influence sensitivity and the proportion of forest core species are unlikely to be an artifact of undersampling in high-disturbance regions (fig. S6). Also,

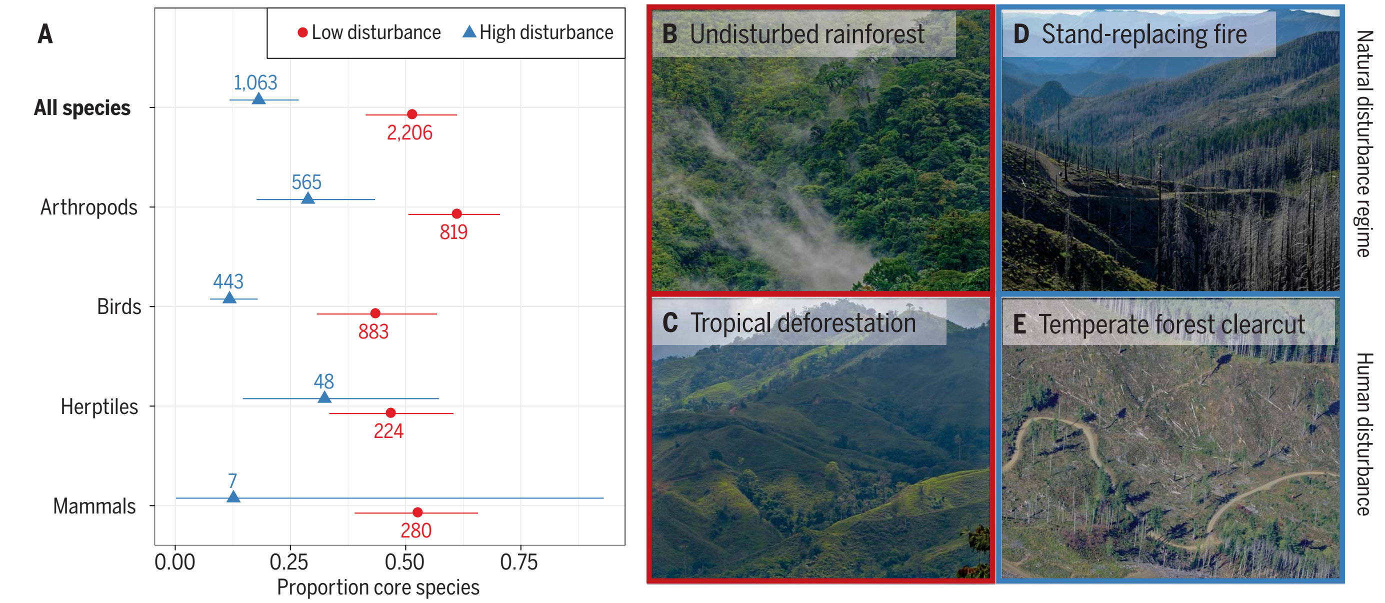


Fig. 2. The proportion of forest species associated with core habitat is mitigated by historical exposure to disturbance. (**A**) Estimated proportions (with 95% confidence intervals) are based on mixed-effects logistic regression models.The binary disturbance variable (low- versus high-disturbance sites) indicates whether each of the 73 BIOFRAG datasets comes from a location that has had high-severity disturbances of any type (glaciation, tropical storms, crown fires, or >50% historical forest loss). Numbers of species are shown beside point estimates. (**B**) Tropical rainforest, undisturbed by stand-replacing disturbance. (**C**) Tropical deforestation for pastureland. (**D**) Temperate forest landscape that has been disturbed by wildfire. (**E**) Temperate forest clearcuts. The extinction filter hypothesis predicts that species in disturbance-prone regions (D) should be less sensitive to habitat edges created by anthropogenic fragmentation (E) than species that have evolved in landscapes where disturbances are rare [(B) and (C)]. [Photos: (B) C. Ziegler; (C) to (E) M. G. Betts]

these results were robust to other potential confounding variables: phylogenetic relatedness (fig. S7), migratory behavior (table S5), geographic range size (table S6), and distance to range edge (table S6). Notably, the strong disturbance effect could not be reproduced when species were categorized using forest amount alone (*6*), indicating that our findings relate primarily to fragmentation in addition to landscape-scale forest loss (tables S7 and S8). The disturbance effect generally remained after statistically accounting for absolute latitude (table S9); the proportion of forest core species declined roughly sixfold and the proportion of forest-associated species declined 1.5-fold over the entire absolute latitudinal gradient observed (0.7° to 51.8°) (Fig. 3 and figs. S8 and S9). Tropical species have been confronted with less historical disturbance (Fig. 1B) and therefore tend to be more edge sensitive and more likely to be associated with forest (table S4 and fig. S8).

The extinction filter hypothesis can be generalized beyond forest species to predict that, in areas typified by large-scale historical disturbances, we should see a greater proportion of species that have evolved with nonforest land-cover types, including disturbed habitats (hereafter, the matrix). For example, a wide range of species in the Pacific Northwestern United States—where stand-replacing crown fires are common—is associated with early successional ecosystems (*22*). Our data support this prediction; the odds of species using matrix habitat relative to using forest habitat were estimated to be 644% higher in high-disturbance regions than in low-disturbance regions (95% credible interval: 523 to 788%) (fig. S5 and table S4). The proportion of matrix species also strongly increased with latitude (fig. S9 and table S4).

Our results support the extinction filter hypothesis; climatic, ecological, and anthropogenic disturbances have already filtered out many of the species that would be more susceptible to forest edges and the process of fragmentation caused by deforestation (*16*). Proportions of forest core species are substantially greater in regions that have not experienced large-scale historical disturbances. This effect results in a latitudinal gradient in fragmentation sensitivity and helps to explain the surprising rarity of extinctions following recent anthropogenic disturbance in Europe and eastern North America (*23*). Species that were strongly sensitive to disturbance-created edges have likely either undergone local extinction or adapted to repeated glaciation or historical land clearance.

Our results partly reconcile the debate about the conservation importance of fragmentation and its effect on biodiversity (*6*, *7*). Many studies that have found reduced fragmentation effects were conducted in already-denuded landscapes (*24*), in locations with stand-replacing disturbance [glaciers or fire (*25*)], and at high latitudes, which experienced glacial advances and retreats (*26*). Conversely, studies that have found strong, negative fragmentation effects are often from the tropics, where broad-scale disturbance is rarer (*27*), or are for matrix-associated temperate zone species, which are adapted to unfragmented but disturbed habitats (*28*, *29*). Exceptions to this general pattern do, of course, exist (*30*, *31*); we caution that temperate species are not necessarily robust to anthropogenic change of other types and synergistic effects of stressors may pose novel threats. For instance, climate change may interact with habitat loss and fragmentation to reduce species' capacity to adapt (*32*). Nevertheless, our data highlight a strong underlying pattern that has the potential to explain why fragmentation studies are known for generating such variable results. It will be essential to tie our broad-scale analyses to the analysis of the mechanistic underpinnings of fragmentation sensitivity to better generalize across biomes and taxa.

These results indicate that conservation actions designed to mitigate edge-driven fragmentation effects can be tailored to the particular regions most likely to host sensitive species, rather than applying simple rules to the entire globe. In regions in temperate zones with greater historical disturbance, efforts might be focused more on conserving mature forest habitat, regardless of its spatial configuration (*6*). On the other hand, efforts to reduce forest fragmentation should be concentrated in regions with low historical

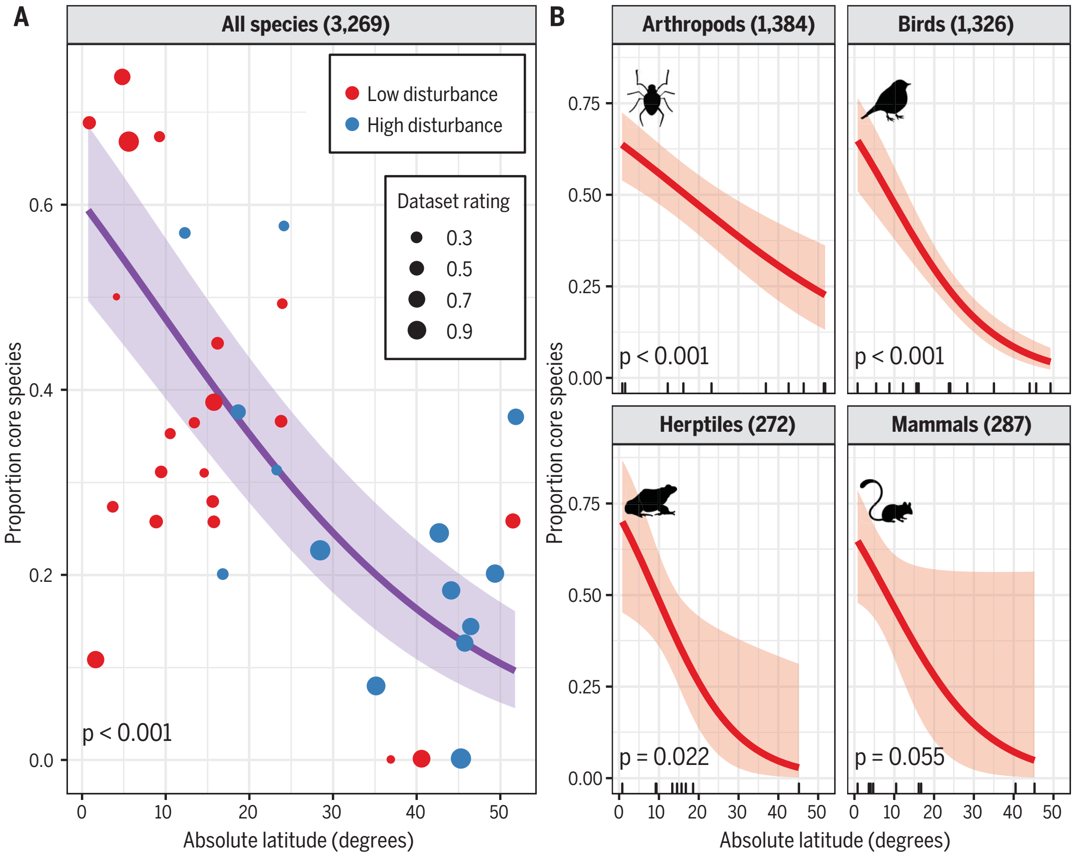


Fig. 3. Logistic regression models used to estimate the proportion of forest core species as a function of absolute latitude. In (**A**), each point shows the proportion of species classified as core within each BIOFRAG region. Point sizes indicate the dataset rating, with higher values reflecting better estimation of edge sensitivity. Point colors indicate disturbance level associated with each region. The response variable is whether or not a species was classified as preferring forest core habitat. Overall, the general pattern observed (decreasing relationship with latitude for forest species) is what one would predict if high-latitude species have evolved to cope with disturbance. Numbers of observations are shown in panel titles. Study region absolute latitudes are shown using black tick marks (**B**). All *P* values were false discovery rate (FDR)–adjusted to control the expected proportion of type I (false positive) errors, and taxonomic class was included as a random effect in the "All species" (A) model. Shaded portions represent 95% confidence bands.

disturbance, particularly tropical forests (*13*)—especially those in biodiversity hotspots (*33*)—where fragmentation continues at a rapid rate and poses the greater extinction risk.

REFERENCES AND NOTES

1. G. Ceballos *et al*., *Sci. Adv.* **1**, e1400253 (2015).
2. T. Newbold *et al*., *Science* **353**, 288–291 (2016).
3. M. G. Betts *et al*., *Nature* **547**, 441–444 (2017).
4. J. M. Diamond, *Biol. Conserv.* **7**, 129–146 (1975).
5. D. S. Simberloff, L. G. Abele, *Science* **191**, 285–286 (1976).
6. L. Fahrig, *Annu. Rev. Ecol. Evol. Syst.* **48**, 1–23 (2017).
7. R. J. Fletcher Jr *et al*., *Biol. Conserv.* **226**, 9–15 (2018).
8. S. B. Carvalho *et al*., *Nat. Ecol. Evol.* **1**, 0151 (2017).
9. N. M. Haddad *et al*., *Sci. Adv.* **1**, e1500052 (2015).
10. F. Taubert *et al*., *Nature* **554**, 519–522 (2018).
11. L. Ries, R. J. Fletcher Jr., J. Battin, T. D. Sisk, *Annu. Rev. Ecol. Evol. Syst.* **35**, 491–522 (2004).
12. L. R. Prugh, K. E. Hodges, A. R. Sinclair, J. S. Brashares, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20770–20775 (2008).
13. M. Pfeifer *et al*., *Nature* **551**, 187–191 (2017).
14. A. Balmford, *Trends Ecol. Evol.* **11**, 193–196 (1996).
15. A. J. Hansen, D. L. Urban, *Landsc. Ecol.* **7**, 163–180 (1992).
16. P. Drapeau, M.-A. Villard, A. Leduc, S. J. Hannon, *Divers. Distrib.* **22**, 385–399 (2016).
17. See supplementary materials.
18. M. Pfeifer *et al*., *Ecol. Evol.* **4**, 1524–1537 (2014).
19. S. Lavorel, M. D. Flannigan, E. F. Lambin, M. C. Scholes, *Mitig. Adapt. Strategies Glob. Change* **12**, 33–53 (2007).
20. N. Ray, J. Adams, *Internet Archaeol.* 10.11141/ia.11.2, (2001).
21. Met Office, Location of tropical cyclones. *Met Office* (2018); https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/hurricanes/location.
22. M. E. Swanson *et al*., *Front. Ecol. Environ.* **9**, 117–125 (2011).
23. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, *Science* **269**, 347–350 (1995).
24. M. K. Trzcinski, L. Fahrig, G. Merriam, *Ecol. Appl.* **9**, 586–593 (1999).
25. K. McGarigal, W. C. McComb, *Ecol. Monogr.* **65**, 235–260 (1995).
26. F. K. Schmiegelow, C. S. Machtans, S. J. Hannon, *Ecology* **78**, 1914–1932 (1997).
27. J. Barlow *et al*., *Nature* **535**, 144–147 (2016).
28. E. I. Damschen, N. M. Haddad, J. L. Orrock, J. J. Tewksbury, D. J. Levey, *Science* **313**, 1284–1286 (2006).
29. I. Hanski *et al*., *Nat. Commun.* **8**, 14504 (2017).
30. R. M. Ewers, S. Thorpe, R. K. Didham, *Ecology* **88**, 96–106 (2007).
31. M. G. Betts, G. J. Forbes, A. W. Diamond, P. D. Taylor, *Ecol. Appl.* **16**, 1076–1089 (2006).
32. J. M. Northrup, J. W. Rivers, Z. Yang, M. G. Betts, *Glob. Change Biol.* **25**, 1561–1575 (2019).
33. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, J. Kent, *Nature* **403**, 853–858 (2000).
34. G. Betts *et al*., Data and Code for "Extinction filters mediate the global effects of habitat fragmentation on animals." v2, Figshare (2019); https://doi.org/10.6084/m9.figshare.9503207.v2.

ACKNOWLEDGMENTS

Funding: Funding from the National Science Foundation (NSFDEB-1457837) and the College of Forestry IWFL Professorship in Forest Biodiversity Research to M.G.B. supported this research. **Author contributions:** M.G.B. and C.W. conceived the study; C.W. and M.G.B. analyzed the data, with original BIOFRAG methods and data curation developed by M.P. and R.M.E.; and M.G.B. and C.W. wrote the first draft of the paper with subsequent review and editing from all other authors, who also contributed datasets to the BIOFRAG database. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data, code, and materials used in the analysis are available from FigShare (*34*) and BIOFRAG (https://biofrag.wordpress.com/biofrag-measuring-biodiversity-response-to-forest-fragmentation/).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1236/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S9
References (*35*–*60*)

6 May 2019; accepted 23 October 2019
10.1126/science.aax9387

ULTRAFAST OPTICS

Ultrafast stimulated emission microscopy of single nanocrystals

Lukasz Piatkowski[1,2]*, Nicolò Accanto[1]†‡, Gaëtan Calbris[1]†, Sotirios Christodoulou[1,3], Iwan Moreels[3,4], Niek F. van Hulst[1,5]*

Single-molecule detection is a powerful method used to distinguish different species and follow time trajectories within the ensemble average. However, such detection capability requires efficient emitters and is prone to photobleaching, and the slow, nanosecond spontaneous emission process only reports on the lowest excited state. We demonstrate direct detection of stimulated emission from individual colloidal nanocrystals at room temperature while simultaneously recording the depleted spontaneous emission, enabling us to trace the carrier population through the entire photocycle. By capturing the femtosecond evolution of the stimulated emission signal, together with the nanosecond fluorescence, we can disentangle the ultrafast charge trajectories in the excited state and determine the populations that experience stimulated emission, spontaneous emission, and excited-state absorption processes.

Complex physical, chemical, and biological processes are determined by fundamental spatial and temporal interaction trajectories. Only ultrafast techniques with single-emitter sensitivity can unveil their inherent transient intermediates and allow exploration of processes such as molecular vibrations and energy transfer (*1*–*3*) and of nanoscale dynamics in plasmonic or two-dimensional materials (*4*, *5*). The small interaction cross sections of individual nanoparticles make it difficult to rely on the conventional ultrafast approaches, such as transient absorption and nonlinear four-wave mixing. Consequently, single molecules and nanoparticles are almost exclusively detected through Stokes-shifted spontaneous emission [fluorescence or photoluminescence (PL)], which is background-free, allowing for photon counting sensitivity and detection of weakly fluorescent emitters. The use of fluorescence detection, however, is hampered by a number of limitations: It is restricted to luminescent samples, is sensitive to bleaching, and, in the linear regime, is slow (nanoseconds), reporting only on the population of the final emitting state and missing out on femtosecond dynamics. Despite the exploration of several alternative detection schemes, such as photothermal (*6*), linear absorption (*7*, *8*), and enhanced Raman (*9*), ultrafast detection of single entities beyond fluorescence has remained challenging.

Here, we demonstrate a highly sensitive experimental scheme based on the direct detection of stimulated emission (SE) for studying the excited-state dynamics in nanoscopic samples with femtosecond temporal resolution. SE microscopy involves one laser pulse for promotion to the excited state and a second, delayed pulse for stimulation back to the ground state, generating a new SE photon (*10*). SE forms the basis of stimulated emission depletion (STED) microscopy; however, in a typical STED experiment, the stimulated photons are discarded and only PL is recorded.

Yet, the instantaneous SE photons contain information on the excited-state population and its dynamics and relaxation pattern, which is otherwise inaccessible from the slow PL. To its advantage, SE is not dependent on the quantum efficiency of the sample, has femtosecond temporal resolution, is coherent, and is capable of mapping the dynamics of an arbitrary excited state.

We present direct stimulated emission detection and imaging of individual nanocrystals (NCs) and trace the excited-state dynamics of single colloidal CdSe/CdS rod-in-rod NCs (*11*) with femtosecond temporal resolution at ambient conditions. The PL is detected simultaneously with the SE, generating two independent, complementary images. It is important to understand the dynamic interplay between various charge relaxation pathways—such as charge injection, extraction, transfer and delocalization, and excited-state relaxation—both ultrafast and with nanoscopic sensitivity (*12*–*14*). Our femtosecond SE experiment on single NCs shows the excited-state relaxation dynamics of individual charges, the dynamical heterogeneity of NCs, and the relative contributions of the various stimulated processes, all with single-NC sensitivity.

A pump beam excites the NC through two-photon absorption to a highly excited state in the conduction band [Fig. 1; for details, see materials and methods (*15*)]. The excited hot electrons and holes, initially localized in the

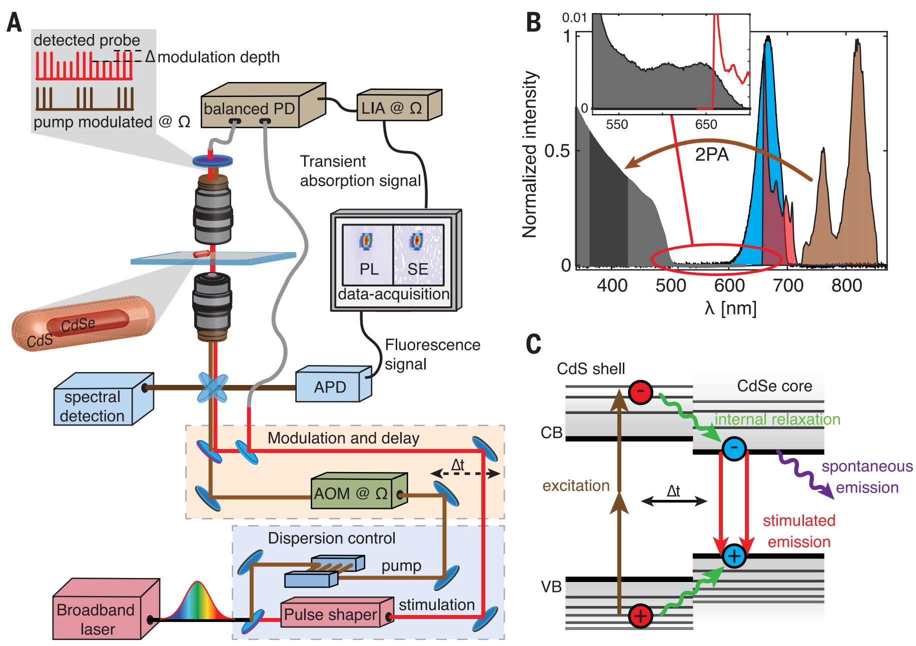


Fig. 1. Concept of the ultrafast stimulated emission nanoscopy. (**A**) Schematic of the experimental setup. PD, photodiode; LIA, lock-in amplifier; APD, avalanche photodiode; AOM, acousto-optic modulator. (**B**) Spectral characteristics of the broadband laser pulse (pump pulse, brown; probe pulse, red) and CdSe/CdS NCs. Gray- and blue-shaded areas represent the absorption and emission spectra of the NCs, respectively. The black area indicates the spectral range of the two-photon absorption (2PA). (**C**) Energy-level sketch of a core/shell CdSe/CdS NC. CB, conduction band; VB, valence band.

[1]ICFO–Institut de Ciences Fotoniques, the Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain. [2]Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland. [3]Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy. [4]Department of Chemistry, Ghent University, Krijgslaan 281-S3, 9000 Gent, Belgium. [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.
*Corresponding author. Email: lukasz.j.piatkowski@put.poznan.pl (L.P.); niek.vanhulst@icfo.eu (N.F.v.H.)
†These authors contributed equally to this work.
‡Present address: Institut de la Vision, Sorbonne Université, Inserm S968, CNRS UMR7210, 17 Rue Moreau, 75012 Paris, France.

shell, decay through the excited-state progression and eventually localize in the lowest excited state (band edge) in the core (Fig. 1C). The probe (stimulation) beam, resonant with the core band-edge transition, leads to charge recombination, stimulates the NC back to the ground state, and induces emission of a stimulated photon. Therefore, any information on the excited charges imprinted by the pump beam in the shell is "read out" by the stimulating probe beam, when one of the excited charges reaches the core band-edge states. The pump beam is modulated, and the SE signal is retrieved by lock-in detection.

As a first step, we raster-scanned the sample while simultaneously detecting both modulated signal (S_{mod}) and PL (Fig. 2, A and B). The PL image clearly reveals the presence of the NCs, which we verified through their emission spectra (fig. S1). The corresponding S_{mod} image shows contrast at the same sample positions where the PL signal appears. Moreover, the measured S_{mod} signal was always positive, meaning we detected extra photons in our stimulation beam (supplementary text 1). Two effects can, in principle, lead to an increase in the transmitted probe-beam intensity when the NC is excited: stimulated emission and ground-state depletion (GSD). In the first case, the SE process following electron-hole recombination gives a net increase in the probe-beam intensity. In the second case, the absorption of the probe beam is lower because of the depletion of the ground state, owing to the presence of either a hole or electron in its respective energy level. The two contributions can be readily distinguished by time-resolved experiments, as shown later. For most NCs, we found a perfect correspondence between PL and S_{mod} images and observed S_{mod} wherever PL appeared (Fig. 2C). Interestingly, in some cases, we detected PL but no measurable S_{mod} (Fig. 2E). We assigned this signal to core-free CdS shell nanoparticles that conucleated during synthesis. Finally, on rare occasions, we observed S_{mod} contrast but no PL (Fig. 2D). The signal likely originated from highly quenched NCs, because it is improbable that we would have observed other particles with the exact same spectral signature. Clearly, the spectral dependence of S_{mod} correlated with the probe beam, and the ability to simultaneously detect PL and S_{mod} gives us extra insight as to the nature of the detected NCs.

Ultrafast coherent response is the main advantage of SE detection. In Fig. 3A, we show a series of PL and S_{mod} images for different interpulse delay times (see fig. S2 for more images). Although the PL signal is detected at all delay times Δt, the S_{mod} signal appears only when the pump pulse overlaps or precedes the stimulation pulse. At negative delay times, when the stimulation pulse arrives before the pump pulse, the NC is in its ground state and there is no excited-state population for the probe pulse to interact with. For the NC marked with an "x" in Fig. 3A, the second-order autocorrelation trace exhibits a dip with degree of coherence $g^{(2)}(0) \lesssim 0.5$, indicating the nonclassical emission of a single NC (fig. S3). The time-resolved traces revealed that when S_{mod} (blue) increases in time, the PL (red) decreases (Fig. 3B). This is intuitive: The excited-state population, which is stimulated down back to the ground state, does not contribute to the spontaneous emission, leading to PL depletion. The fact that S_{mod} and PL signals are anticorrelated unambiguously indicates that S_{mod} contains a substantial contribution from the SE process. Furthermore, we found that the changes in both signals, S_{mod} ingrowth (ΔS_{mod}) and PL depletion (ΔPL), occur on specific time scales. Interestingly, the ΔPL depletion occurs with a single time constant, whereas ΔS_{mod} grows in with two time constants. The slower time constant of ~400 to 700 fs is identical to the time constant with which ΔPL decreases. However, a considerable part of the S_{mod} grows on a faster time scale and cannot be observed within the cross-correlation time of the pump and probe pulses (<200 fs). To understand this, one needs to consider that the NCs are initially pumped to a highly excited state in the shell (supplementary text 3), whereas the stimulation pulse probes the lowest excited state in the core. GSD occurs when charges are present in the excited state of the transition resonant with the probe energy. As soon as the faster of the two charges reaches the lowest excited state of the core (*16*–*19*), the probe-beam absorption will decrease. This means that GSD reports on the relaxation rate of the fastest charge, either the electron or the hole. By contrast, the probe beam can induce charge recombination and SE only when both electron and hole localize into the core. Consequently, SE is sensitive to relaxation of the slower of the two charges. In the PL, we see only the slower component because PL is a time-averaged signal, which is mostly sensitive to the population decay of the lowest excited state (supplementary text 4).

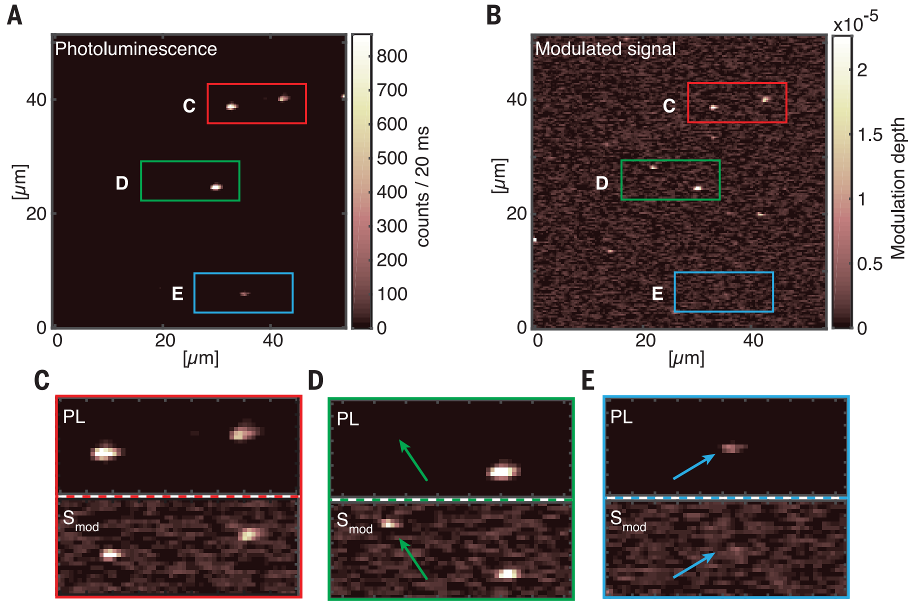


Fig. 2. Stimulated emission imaging. (**A** and **B**) Confocal images of the same sample area showing PL and the lock-in signal (S_{mod}), respectively. The stimulation beam was set to arrive 7 ps after the pump beam (supplementary text 2). (**C** to **E**) Comparison between the PL and S_{mod} images for the three regions of interest indicated in (A) and (B).

We quantified the observed dynamics by simultaneously fitting the PL and S_{mod} traces (supplementary text 5). PL and S_{mod} traces acquired on small NC clusters revealed that the average slower charge relaxation time is 550 fs (black histogram in Fig. 3C). The time-delay traces recorded repeatedly on the same individual NCs (for more traces, see fig. S5) revealed the relaxation heterogeneity among individual NCs (Fig. 3C). From the difference in the dynamics between SE and GSD, we determined the relative contribution of the two signals to the total measured signal S_{mod} by performing simple, qualitative kinetic rate equation calculations (supplementary text 6). The experimental ratio of SE/S_{mod} extracted from individual time traces for a large number of NCs centers around a value of ~0.17 (Fig. 3D). The observation of a ratio SE/S_{mod} < 0.2 strongly suggests that the cross sections for absorption and SE might be somewhat different, given the large asymmetry between the shape of the absorption and emission bands.

The lower SE signal with respect to GSD signal might also be caused by an excited-state

absorption (ESA) process. In ESA, the probe beam promotes the excited charges to higher excited states at the cost of absorbing a probe-beam photon, leading to a reduction of the apparent SE contrast and enhanced bleaching (*14*, *20*) and quenching (*21*). To uncover the role of ESA in our NC dynamics, we varied the duration of the probe pulse, because the ESA timing should be sensitive to the observed 550-fs relaxation time of the hot state. Once the charges have again returned to the emitting state, the probe pulse should stimulate the NC down. The concept, depicted in Fig. 4B, is analogous to STED experiments, where the STED pulse is stretched to prevent reexcitation (*22*). We measured the S_{mod} and ΔPL contrast for increasing probe-pulse duration (Γ), stretched up to 2.5 ps, at Δt = 7 ps delay. In Fig. 4C, both S_{mod} and ΔPL show increased contrast with the probe-pulse duration. Interestingly, the ingrowth matches very well the 550-fs excited-state charge relaxation time determined from the pump-probe traces. A simulation using the kinetic rate equation model expanded with the ESA process (supplementary text 7) reproduces the experimental data well and confirms our hypothesis that stretching the stimulating probe pulse allows stimulation down of charges that otherwise undergo ESA.

Interestingly, the simultaneous detection of stimulated and spontaneous emission of a single NC allows us to correlate the decays in a quantitative manner. The number of photons detected in SE should be equal to the number of photons missing in PL, that is, PL depletion. For the data shown in Fig. 3B, we determined an effective number of photons depleted from PL, ΔPL_{eff} = 1.6 × 10^7 photons/s, and an effective number of photons gained in the stimulation beam, ΔSE_{eff} = 1.3 × 10^7 photons/s per NC (supplementary text 8). The two values are in good agreement, given that the detection occurs in two independent channels, using photon counting versus analog detectors.

The high sensitivity of the presented SE detection opens up new imaging possibilities for weakly fluorescing or quenched systems, and the time-resolved experiment provides information on the excited-state relaxation dynamics and its mechanism, all with femtosecond time resolution and single-emitter sensitivity. The unconventional, simultaneous detection of the spontaneous and stimulated emission provides large imaging specificity: The fact that SE depends on two distinct frequencies, in combination with the interpulse time delay, makes the method extremely

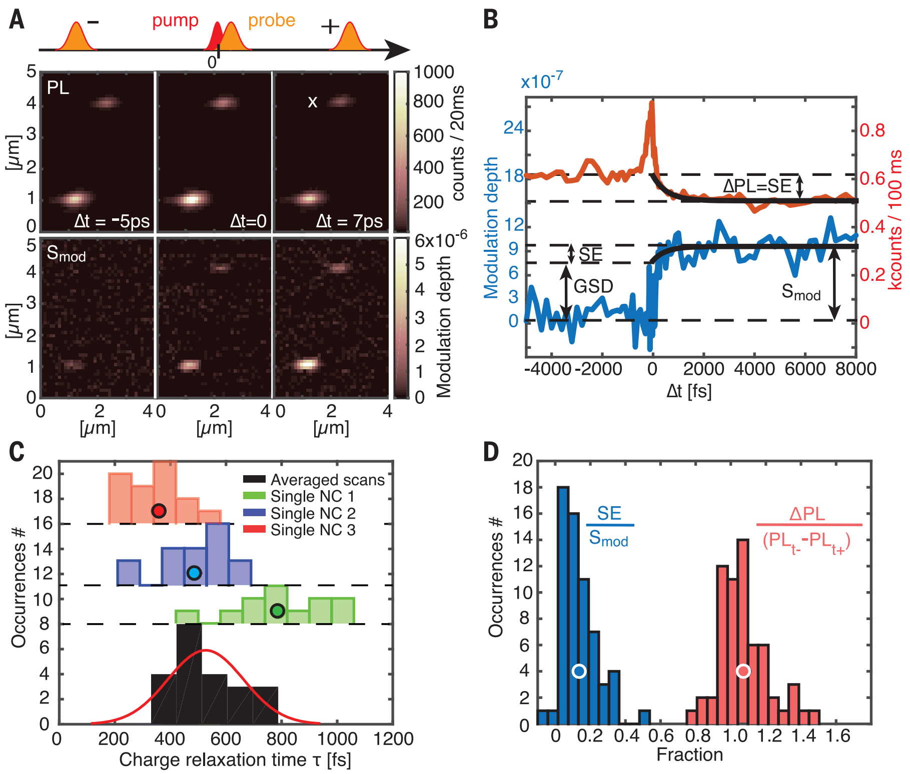


Fig. 3. Time-resolved stimulated emission microscopy. (**A**) A series of images acquired by detecting PL and the S_{mod} signal for different excitation and stimulation interpulse delays Δt. (**B**) Simultaneously detected S_{mod} (blue) and PL (red) time traces for a CdSe/CdS NC. (**C**) Histogram of the exciton relaxation times. Red, blue, and green histograms correspond to relaxation times extracted from the fits to individual time traces of three different, single NCs. The black histogram shows occurrences of relaxation times extracted from averaged traces from NC clusters. (**D**) Histograms showing the relative contributions of the SE (blue) and the ΔPL (red) to the total detected signal change S_{mod} and $PL_{t-} - PL_{t+}$, respectively.

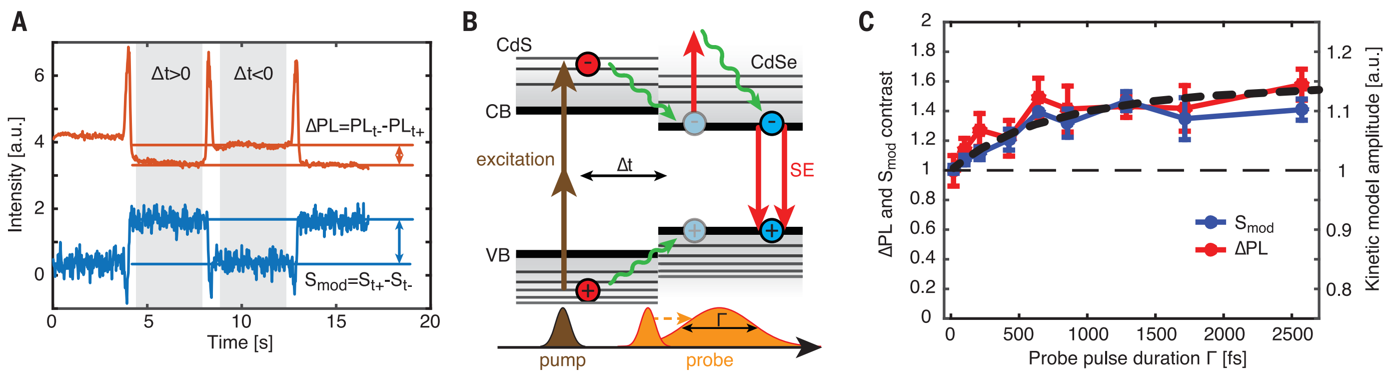


Fig. 4. Higher stimulated emission and photoluminescence contrast with longer probe pulse. (**A**) PL and S_{mod} signals recorded in time while repeatedly scanning the interpulse delay time Δt from negative to positive values. a.u., arbitrary units. (**B**) Concept of the varying probe-pulse duration experiment. (**C**) Normalized S_{mod} and ΔPL as a function of probe-pulse duration. The traces were averaged from seven separate measurements (four positively and three negatively chirped probe traces) on different NC clusters. Error bars indicate the standard deviation. The black dashed line is the result of solving the set of kinetic rate equations described in supplementary text 7.

sensitive to different species within a dense ensemble.

The time-resolved femtosecond SE experiment allowed to us to provide a comprehensive picture of the excited charges, which are either stimulated down or promoted to higher excited states or recombine spontaneously. The SE and GSD contributions comprise <20 and >80% of the total induced ground- and excited-state population difference, respectively. This was aided by the fact that the two excited charges—electrons and holes—exhibit different relaxation times (supplementary text 9). The rod-in-rod CdSe/CdS NC excited holes localize at the core band edge within 200 fs, whereas the excited electrons relax to the core band edge on a time scale of 550 fs. We found that the electron relaxation time differs by nearly a factor of two between individual NCs. Finally, the single-emitter sensitivity of our experiment allowed us to compare the number of photons lost in PL and gained through SE in absolute terms, which is difficult to achieve for ensembles (*23*). Stretching the stimulation pulse allowed us to elucidate the presence of ESA and increase the SE efficiency by 40 to 50%, that is, a substantial portion of the excited charges undergo ESA and relax back to the core band-edge states.

The ultrafast SE microscopy opens up a spectrum of experiments for exploration (supplementary text 10). Scanning the stimulation pulse energy would allow for state selectivity and enable the study of excited state-to-state dynamics (*16*). Given its coherent nature, SE microscopy could be expanded to accommodate heterodyne detection of the stimulation beam and could provide easy access to investigating coherent effects such as coherent energy transfer (*3*, *24*). The absorption cross section of our NCs at the stimulation wavelength is approximately an order of magnitude larger than the absorption cross section of a typical fluorescent dye (3×10^{-16} versus 10^{-17} cm^2) (*25*). Therefore, even single molecules could be detected in stimulated emission.

REFERENCES AND NOTES

1. S. Yampolsky *et al.*, *Nat. Photonics* **8**, 650–656 (2014).
2. M. Liebel, C. Toninelli, N. F. van Hulst, *Nat. Photonics* **12**, 45–49 (2018).
3. R. Hildner, D. Brinks, J. B. Nieder, R. J. Cogdell, N. F. van Hulst, *Science* **340**, 1448–1451 (2013).
4. M. Aeschlimann *et al.*, *Science* **333**, 1723–1726 (2011).
5. V. Kravtsov, R. Ulbricht, J. M. Atkin, M. B. Raschke, *Nat. Nanotechnol.* **11**, 459–464 (2016).
6. A. Gaiduk, M. Yorulmaz, P. V. Ruijgrok, M. Orrit, *Science* **330**, 353–356 (2010).
7. P. Kukura, M. Celebrano, A. Renn, V. Sandoghdar, *J. Phys. Chem. Lett.* **1**, 3323–3327 (2010).
8. S. Chong, W. Min, X. S. Xie, *J. Phys. Chem. Lett.* **1**, 3316–3322 (2010).
9. A. B. Zrimsek *et al.*, *Chem. Rev.* **117**, 7583–7613 (2017).
10. W. Min *et al.*, *Nature* **461**, 1105–1109 (2009).
11. S. Christodoulou *et al.*, *Nat. Commun.* **6**, 7905–7913 (2015).
12. J. Hanne *et al.*, *Nat. Commun.* **6**, 7127–7133 (2015).
13. M. D. Lesoine *et al.*, *J. Phys. Chem. C* **117**, 3662–3667 (2013).
14. S. E. Irvine, T. Staudt, E. Rittweger, J. Engelhardt, S. W. Hell, *Angew. Chem. Int. Ed.* **47**, 2685–2688 (2008).
15. See the supplementary materials.
16. P. Kambhampati, *J. Phys. Chem. C* **115**, 22089–22109 (2011).
17. E. Hendry *et al.*, *Phys. Rev. Lett.* **96**, 057408–057412 (2006).
18. S. Brovelli *et al.*, *Nano Lett.* **14**, 486–494 (2014).
19. M. Zavelani-Rossi, M. G. Lupo, F. Tassone, L. Manna, G. Lanzani, *Nano Lett.* **10**, 3142–3150 (2010).
20. J. Hotta *et al.*, *J. Am. Chem. Soc.* **132**, 5021–5023 (2010).
21. T. Watanabe *et al.*, *Chem. Phys. Lett.* **420**, 410–415 (2006).
22. T. A. Klar, S. W. Hell, *Opt. Lett.* **24**, 954–956 (1999).
23. E. Rittweger, B. R. Rankin, V. Westphal, S. W. Hell, *Chem. Phys. Lett.* **442**, 483–487 (2007).
24. A. Chenu, G. D. Scholes, *Annu. Rev. Phys. Chem.* **66**, 69–96 (2015).
25. L. Kastrup, S. W. Hell, *Angew. Chem. Int. Ed.* **43**, 6646–6649 (2004).

ACKNOWLEDGMENTS

L.P. acknowledges the Marie Skłodowska-Curie COFUND and the ICFOnest programs. **Funding:** This project received funding from the National Science Centre, Poland, grant 2015/19/P/ST4/03635, POLONEZ 1, and from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 665778. This research was funded by the European Commission (ERC Advanced Grant 670949-LightNet), the Spanish Ministry of Economy MINECO (FIS2012-35527, FIS2015-72409-EXP, FIS2015-69258-P, Network FIS2016-81740-REDC "NanoLight," and Severo Ochoa Grant SEV2015-0522), the Catalan AGAUR (no. 2017SGR1369), Fundació Privada Cellex, Fundació Privada Mir-Puig, and Generalitat de Catalunya through the CERCA Program. **Author contributions:** L.P. and N.F.v.H. designed the experiment. L.P., N.A., and G.C. performed the experiments and data analysis. S.C. and I.M. provided the samples. L.P. and N.F.v.H. wrote the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1240/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
References (*26*–*35*)

30 May 2019; accepted 7 November 2019
10.1126/science.aay1821

MESOSCOPIC PHYSICS

Transmitting the quantum state of electrons across a metallic island with Coulomb interaction

H. Duprez[1]*, E. Sivre[1]*, A. Anthore[1,2], A. Aassime[1], A. Cavanna[1], U. Gennser[1], F. Pierre[1]†

The Coulomb interaction generally limits the quantum propagation of electrons. However, it can also provide a mechanism to transfer their quantum state over larger distances. Here, we demonstrate such a form of electron teleportation across a metallic island. This effect originates from the low-temperature freezing of the island's charge Q which, in the presence of a single connected electronic channel, enforces a one-to-one correspondence between incoming and outgoing electrons. Such faithful quantum state imprinting is established between well-separated injection and emission locations and evidenced through two-path interferences in the integer quantum Hall regime. The additional quantum phase of $2\pi Q/e$, where e is the electron charge, may allow for decoherence-free entanglement of propagating electrons, and notably of flying qubits.

A disordered environment, with a large number of interacting degrees of freedom, is generally considered unfavorable for quantum technologies. Such an environment is exemplified by a metallic island with a large energy density of states and a small number of connected electronic channels, through which there is no quantum coherent propagation of individual electrons. Indeed, the time that an individual electron spends inside the island (*1*) is typically much longer than the interval between inelastic collisions destroying its quantum coherence (*2*, *3*). In contrast to this conventional wisdom, we show experimentally that the Coulomb interaction in such an island can, under the right circumstances, lead to a near perfect transmission of the quantum state of electrons across it, mediated by the collective surface plasmon modes of the island (*4*, *5*). In the quantum Hall regime implementation, where injection and emission points are physically separated by chirality, this constitutes a form of teleportation of the electrons' state. This phenomenon is different from the standard "quantum teleportation" protocol (*6*) and similar to the "electron teleportation" proposed in (*7*).

The voltage probe model of a metallic island (*8*) is widely used to mimic the electrons' quantum decoherence and energy relaxation toward equilibrium [see, e.g., (*9*)]. However, independent absorption and emission of electrons result in fluctuations of the total island charge Q, with a characteristic charging energy $E_C = e^2/2C$ (C is the geometrical capacitance of the island and e is the elementary electron charge). At low temperatures $T \ll E_C/k_B$ (where k_B is the Boltzmann constant), this energy is not available, and the macroscopic quantum charge state Q is effectively frozen (*5*, *10*) [although not quantized in units of e as long as one channel is perfectly connected (*11*–*13*)]. Consequently, correlations develop between absorbed and emitted electrons. Such correlations are strongest if only one transport channel is connected to the island, in which case theory predicts that the electrons entering it and those concomitantly exiting it are in identical quantum states (*4*, *5*) [see also (*14*) for a related prediction in the presence of strong nonlocal interactions along quantum Hall edges]. Effectively, the electronic states within the connected quantum channel are decoupled from the many quasiparticles within the island, even though the incoming (outgoing) physical electron particles penetrate into (originate from) the island. Another consequence is that heat evacuation from the island's internal states along the channel is fully suppressed (*10*). By contrast, in the presence of two or more open channels, the coherence is lost (*5*), and heat evacuation is restored, in agreement with the recently observed systematic heat Coulomb blockade of one ballistic channel (*15*). The comparable "electron teleportation" proposed in (*7*) also relies on the "all-important" Coulomb charging energy of a small island, which has to prevail over temperature and voltage bias. In this proposal (*7*), the island is superconducting, without subgap states except for Majorana bound states at the injection and emission locations. Such a teleportation process was recently invoked as one possible mechanism for the observed coherent single-electron transport across a hybrid superconductor-semiconductor island in the Coulomb blockade regime (*16*). We additionally point out some similarities with quasiparticle correlations induced by Andreev reflections that take place at the interface between a normal metal and a superconductor (*2*). These correlations can be created nonlocally, through the so-called crossed Andreev reflection involving an electron and a hole separated by at most the superconducting coherence length.

We demonstrate the high-fidelity replication of the quantum state of electrons across a metallic island through quantum interferences. For this purpose, an injected current is first split along two separate paths that are subsequently recombined, thereby realizing an electronic Mach-Zehnder interferometer (MZI).

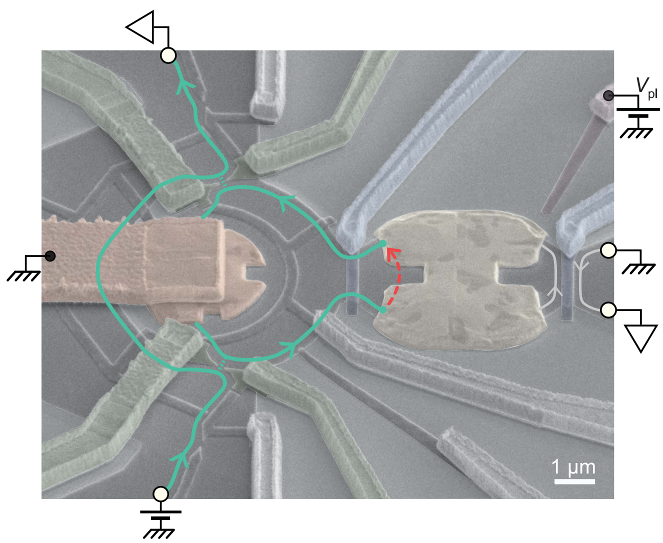


Fig. 1. Device e-beam micrograph. Areas with a Ga(Al)As 2DEG underneath the surface appear darker. The applied perpendicular magnetic field $B \simeq 5$ T corresponds to the integer quantum Hall regime at a filling factor of 2. Capacitively coupled gates colored green and blue control, respectively, the MZI beam splitters for the outer quantum Hall edge channel (lines with arrow, here corresponding to the schematic in Fig. 2B) and the connection to the floating metallic island (sand yellow, in right half) in good ohmic contact with the buried 2D electron gas. One of the two MZI outputs is the central small ohmic contact (light brown, in left half) connected to ground through a suspended bridge. The second one, larger and located farther away, is schematically represented by the top white circle. The MZI phase difference is controlled through B or the plunger gate voltage V_{pl}. The red dashed line visually represents the nonlocal quantum state transfer across the island, between electrons' injection (starting point) and emission (arrow).

[1]Université Paris-Saclay, CNRS, Centre de Nanosciences et de Nanotechnologies (C2N), 91120 Palaiseau, France [2]Université de Paris, C2N, 91120 Palaiseau, France
*These authors contributed equally to this work.
†Corresponding author. Email: frederic.pierre@c2n.upsaclay.fr

In contrast with usual MZI implementations (*17*–*21*), one of the paths can controllably be diverted toward a small floating metallic island (Fig. 1). In that case, any two-path quantum interferences involve both the initial electrons (direct left path) and the reemitted ones (interrupted right path, assuming a perfect contact with the island). Therefore high-visibility interferences directly ascertain a high fidelity of the electron state replication. The Coulomb interaction was previously invoked to account for various observations in electron interferometers, such as the multiple side lobes versus voltage bias in (*22*) or the phase rigidity versus magnetic field in (*23*).

A colorized e-beam micrograph of the measured device is shown in Fig. 1. The sample was nanofabricated from a high-mobility Ga(Al)As two-dimensional electron gas (2DEG) and immersed in a perpendicular magnetic field $B \simeq 5$ T corresponding to the integer quantum Hall filling factor $\nu = 2$, where the electron quantum coherence length is the largest (*18*, *19*, *21*). In this regime, two quantum Hall channels copropagate along the edges (the electron gas was etched away in the brighter areas), and the MZI is formed using only the outer edge channel. The followed paths are represented by thick lines with arrows for the configuration where one MZI arm goes through the floating metallic island (corresponding schematic shown in Fig. 2B). The two MZI beam splitters, each tuned to half transmission, are realized with quantum point contacts formed by field effect using split gates (colored green; the inner quantum Hall channel, not shown, is fully reflected). One of the two MZI outputs is the small central metallic electrode (light brown, in left half of Fig. 1), which is grounded through a suspended bridge. The quantum interferences are characterized by the oscillations of the current transmitted to the second MZI output formed by a much larger electrode 60 μm away (represented in Fig. 1 by the top white circle), while sweeping either the magnetic field B or the voltage V_{pl} applied to a lateral plunger gate (purple). The floating metallic island (sand yellow, in right half of Fig. 1) consists of 2 $\mu\mathrm{m}^3$ of a gold-germanium-nickel alloy diffused into the Ga(Al)As heterojunction by thermal annealing. The typical metallic density of states of such metals is $\nu_{\mathrm{F}} \approx 10^{47}\ \mathrm{J}^{-1}\,\mathrm{m}^{-3}$ (1.14×10^{47} for gold, the main constituent), corresponding to a very small average electronic level spacing in the island of $\delta \approx 30$ peV. The dwell (wandering) time of individual electrons within this island is given by the expression $\tau_{\mathrm{D}} = \frac{h}{N\delta} \approx \frac{130}{N}$ μs, where h is the Planck constant and N is the number of connected edge channels (*1*). As pointed out in the introduction, this is much longer than the quantum coherence time of electrons in similar metals, which is at most in the 20-ns range (*3*, *24*). The gates barring the broad way on each side of the floating island (blue) are normally tuned to either fully reflect or fully transmit the outer edge channel, in order to implement the MZI configurations schematically represented in Fig. 2, A to C. The second (inner) quantum Hall edge channel is always completely reflected at the barring gate (fig. S1) and can therefore be ignored (*5*). The island charging energy $E_{\mathrm{C}} \simeq k_{\mathrm{B}} \times 0.3\,\mathrm{K}$ was obtained from standard Coulomb diamond measurements [in a specifically tuned tunnel regime; see Fig. 3B and (*25*)]. At the experimental electronic temperature $T \simeq 10$ mK [measured on-chip from shot noise (*26*)], the criterion $k_{\mathrm{B}}T \ll E_{\mathrm{C}}$ for fully developed Coulomb-induced correlations is therefore well verified. The previous experiments were performed in the opposite "high-temperature" regime $k_{\mathrm{B}}T \gg E_{\mathrm{C}}$ of negligible Coulomb correlations, in which case, unsurprisingly, a complete quantum decoherence (*27*) and energy relaxation (*28*) of electrons were observed with a single connected channel. Finally, the transparency of the contact between the floating island and the outer quantum Hall edge channel plays an essential role: If the transparency is poor, many electrons would simply be reflected at the interface. Here, $\gtrsim 97\%$ of the incoming current penetrates into the floating island (*25*), which is also reflected by the marked changes of behavior detailed later.

In Fig. 2D, we show typical MZI oscillations versus B of τ_{MZI}, the fraction of outer edge channel current transmitted across the device. The measurements were performed in the three configurations depicted in Fig. 2, A to C. The red continuous line in Fig. 2D corresponds to a standard electronic MZI, with the floating metallic island bypassed (schematic in Fig. 2A). In that case, the oscillations are of high visibility $\mathcal{V} \equiv (\tau_{\mathrm{MZI}}^{\max} - \tau_{\mathrm{MZI}}^{\min})/(\tau_{\mathrm{MZI}}^{\max} + \tau_{\mathrm{MZI}}^{\min}) \approx$ 90% and, as expected for the Aharonov-Bohm phase, the magnetic field period of 241 ± 3 μT (red symbols in Fig. 2E show consecutive extrema positions) closely corresponds to one flux quantum ($241\ \mu\mathrm{T} \times S \simeq 0.98h/e$ using the nominal area $S \simeq 16.8\ \mu\mathrm{m}^2$). A small asymmetry in the τ_{MZI} data (the average is slightly above 0.5) results from a small reflection of the outer edge channel on the grounded central ohmic contact [of ≈5%; see (*25*)]. The black continuous line in Fig. 2D was measured with the right MZI arm deviated to go through the floating ohmic island (edge channel paths displayed in Fig. 1, and schematic in Fig. 2B). We observe first that the visibility of quantum interference remains of the same high amplitude, which corresponds to a perfect fidelity (at experimental accuracy) of the replicated quantum states imprinted on the electrons reemitted from the island, in agreement with low-temperature predictions (*4*, *5*). Second, the magnetic field period of 305 ± 4 μT is found

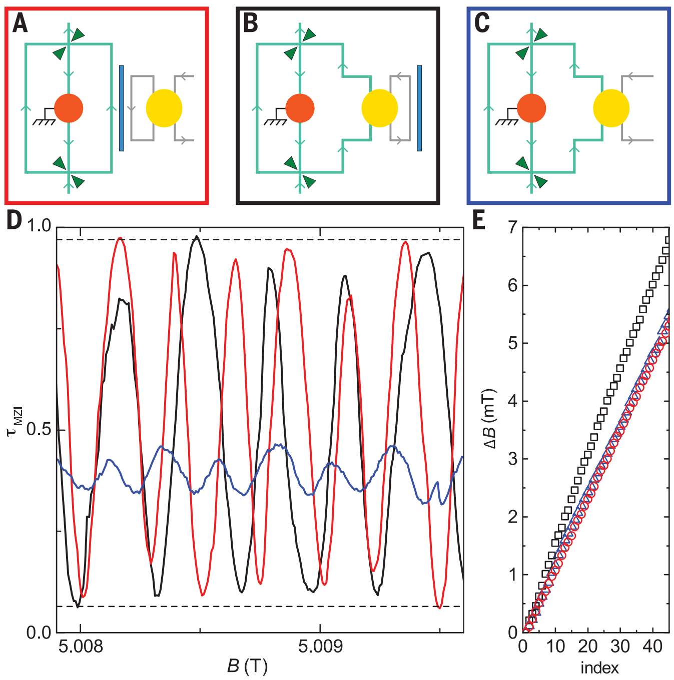


Fig. 2. Quantum oscillations versus magnetic field. (**A** to **C**) Schematics of implemented MZI configurations. (**D**) Fraction τ_{MZI} of the outer edge channel current transmitted across the MZI as a function of B. Continuous lines are measurements performed in the configuration framed by a box of the same color in (A), (B), and (C). The horizontal black dashed lines represent the τ_{MZI} extrema for the standard and floating island MZI configurations [schematics in (A) and (B), respectively], corresponding to a high quantum oscillations visibility of $\mathcal{V} \sim 90\%$. With a second channel connected to the floating island (configuration shown Fig. 2C), the quantum oscillations are strongly reduced to a visibility $\mathcal{V} \sim 20\%$, consistent with the separately characterized small residual reflection of ~3% [see text and (*25*)], and the average τ_{MZI} is diminished as part of the current is transmitted across the island toward a remote electrical ground. (**E**) Symbols display the magnetic field position of consecutive extrema (both peaks and dips increment the index number).

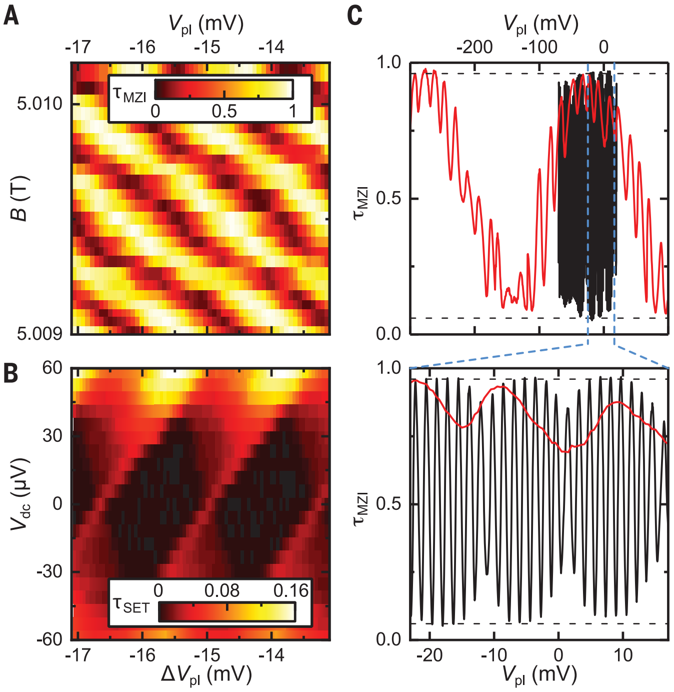


Fig. 3. Quantum phase versus island charge. (**A**) Color plot of $\tau_{MZI}(B, V_{pl})$ in the floating island MZI configuration (schematic in Fig. 2B), with the larger values shown brighter, which establishes the equivalent role of B and V_{pl}. (**B**) Coulomb diamond characterization of the floating island (larger differential conductance τ_{SET} shown brighter; in this configuration, the island is weakly coupled on both sides and V_{dc} is the applied dc bias voltage). A comparison with (A), plotted using the same V_{pl} scale, reveals that the addition of a charge of e on the island precisely corresponds, in the floating island MZI configuration, to an electron quantum phase of 2π (one quantum oscillation period). (**C**) The top and bottom panels display measurements of $\tau_{MZI}(V_{pl})$ with the device set in the floating island MZI configuration (black line) and in the standard MZI configuration (red line, schematic in Fig. 2A). The MZI oscillations' period in V_{pl} is shorter by a factor of 1/160 when the island is connected. An additional modulation of fixed period (≈15 mV) appears in both configurations.

to be larger than in the standard MZI configuration of Fig. 2A (see black symbols in Fig. 2E). This increase is opposite to the reduction that would be expected from the Aharonov-Bohm period with the larger surface enclosed by the outer channel path and the inner boundary of the floating metallic island [see (*25*) for a graphical representation; $S \simeq 18.4\ \mu\text{m}^2$ would correspond to an Aharonov-Bohm period of $225\ \mu\text{T} \simeq h/eS$]. Such opposite evolution and relatively important discrepancy (36%) establish that the MZI phase does not reduce to the usual Aharonov-Bohm phase acquired by a single electron propagating along two different edge paths. Even a naïve application of the surface modulation that cancels the magnetic field dependence in Fabry-Pérot interferometers dominated by Coulomb interaction (*29*, *30*) would only compensate for the period reduction by the added area enclosed between the metallic island and the barring gate. Instead, the larger period corroborates the transfer of the electrons' state across the island. Indeed, in the presence of an electron path amputated from a section (the 2DEG/metal interface), the closed surface involved in the Aharonov-Bohm phase is no longer well-defined. Whether one can still speak of an Aharonov-Bohm phase with a smaller effective surface, or whether another period reduction mechanism is at play when going through the floating island, is not established. This question calls for further investigations, both theoretical and experimental, with devices implementing different injection-emission distances.

The blue continuous line in Fig. 2D was measured with one MZI arm going through the floating island and in the presence of a second electronic channel connected to it (configuration schematically displayed in Fig. 2C). We find strongly suppressed conductance oscillations corresponding to a full decoherence of the electrons going through the island. The residual visibility $\mathcal{V} \lesssim 0.2$ is consistent with the separately obtained proportion $1 - \tau_{island} \lesssim 3\%$ of reflected electrons, not penetrating into the island (*25*). Indeed, the MZI contribution of the reflected electrons reads $\mathcal{V}(1 - \tau_{island} \ll 1) \simeq \left(\frac{4\mathcal{V}_0}{3}\right)\sqrt{1 - \tau_{island}} \lesssim 0.21$, with $\mathcal{V}_0 \approx 90\%$ the MZI visibility in the standard configuration (*25*, *27*). The magnetic field period of $246 \pm 4\ \mu\text{T}$ for these smaller oscillations [see blue symbols in Fig. 2E; a standard fast Fourier transform analysis of these residual oscillations yields the less resolved value $237 \pm 16\ \mu\text{T}$ (*25*)] is 9% longer than the 225 μT nominally expected, and is close to the period observed in the standard MZI configuration shown in Fig. 2A. It is plausible that the reflected electrons are those propagating the furthest away from the edge and from the semiconductor-metal interface, effectively corresponding to a smaller surface compatible with the observed longer period. Other possibilities include that the residual electrons' reflection takes place at the level of the barring gate (colored blue, left of island in Fig. 1) or, eventually, a Coulomb-induced compensation in the spirit of Coulomb dominated Fabry-Pérot interferometers (*29*, *30*). Also, the average $\langle \tau_{MZI} \rangle \simeq 0.39$ is shifted below 0.5. This is simply because part of the injected current is evacuated toward a remote electrical ground through the second (right) channel connected to the floating island (the value $\langle \tau_{MZI} \rangle = 0.375$ is expected from current conservation for a floating island and a central ohmic contact both perfectly connected).

We now investigate the relation between the island's charge and the electron phase shift associated with the quantum state transfer. For this purpose, Fig. 3 focuses on the influence on τ_{MZI} of the voltage V_{pl} applied to a plunger gate (colored purple in Fig. 1), which is relatively far from the MZI outer quantum Hall channel but close to the island. The equivalent role on the MZI phase of V_{pl} and B is first directly established, in Fig. 3A, with the device set in the floating island MZI configuration [schematic in Fig. 2B, see also, e.g., (*31*, *32*) for the influence of an electrostatic field on quantum interferences through different mechanisms]. Figure 3B displays Coulomb diamond measurements of the conductance across the island as a function of the same plunger gate voltage V_{pl}. For this purpose, the island is here weakly connected through tunnel barriers, thereby implementing a single-electron transistor of quantized charge Q in units of e (Q is not quantized for the strongly coupled floating island in the MZI configurations). The left MZI arm was disconnected during this measurement, as schematically illustrated in fig. S2 (*25*). Notably, the MZI gate voltage period in Fig. 3A precisely matches the Coulomb diamonds' period in Fig. 3B, as can be seen by directly comparing the two panels plotted using the same V_{pl} scale. In the floating MZI limit of strongly connected channels, $Q = eV_{pl}/\Delta$, with $\Delta \simeq 1.7$ mV being the Coulomb diamond period (*11*–*13*). A quantum phase shift of $2\pi Q/e$ therefore applies to the transferred electrons, as specifically predicted theoretically (*4*, *5*) and in agreement with Friedel's sum rule. Figure 3C shows a comparison of τ_{MZI} oscillations in the standard MZI configuration (red line) and with one arm going through the metallic island (black line); both have maximum visibility $\mathcal{V} \simeq 90\%$ as also seen versus magnetic field in Fig. 2D. However, the V_{pl} period is increased by a large factor of 160, from 1.7 to 270 mV, when the island is disconnected; this reflects the weak coupling of the plunger gate voltage to the MZI outer edge channel [see fig. S3 (*25*) for an

extended V_{pl} range]. This provides further evidence that the edge electrons contributing to the quantum oscillations in the floating island configuration are indeed incorporated into the metal. An additional, smaller signal of fixed period 15 mV is visible in both configurations (in the form of direct oscillations or of an amplitude modulation), which might originate from the progressive charging of a nearby defect.

This experimental work demonstrates that the Coulomb interaction has two facets. It can both destroy and preserve quantum effects. Although a metallic island is often pictured as a floating reservoir of uncorrelated electrons (*8*, *33*), we establish that a high-fidelity electron quantum state transfer can take place across it, enforced by the Coulomb charging energy. This provides a means to overcome limitations imposed by the decoherence of individual electrons. Moreover, the observed universal 2π electron phase shift for one elementary charge e on the island can allow for a strong entanglement of single-electron states, both between themselves or with other quantum degrees of freedom, with a negligible loss of coherence. Such controllable, strong-coupling mechanism constitutes a key element in the context of quantum Hall edges envisioned as platforms for the manipulation and transfer of quantum information via propagating electrons (*21*, *34*–*39*). In particular, it is very well suited to implementing quantum gates for these “flying qubits,” such as the CNOT proposal involving a conditional phase shift of π described in (*38*).

REFERENCES AND NOTES

1. P. Brouwer, M. Büttiker, *Europhys. Lett.* **37**, 441–446 (1997).
2. Y. Nazarov, Y. Blanter, *Quantum Transport* (Cambridge Univ. Press, 2009).
3. F. Pierre *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **68**, 085413 (2003).
4. A. Clerk, P. Brouwer, V. Ambegaokar, *Phys. Rev. Lett.* **87**, 186801 (2001).
5. E. G. Idrisov, I. P. Levkivskyi, E. V. Sukhorukov, *Phys. Rev. Lett.* **121**, 026802 (2018).
6. C. H. Bennett *et al.*, *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
7. L. Fu, *Phys. Rev. Lett.* **104**, 056402 (2010).
8. M. Büttiker, *IBM J. Res. Develop.* **32**, 63–75 (1988).
9. Y. M. Blanter, M. Büttiker, *Phys. Rep.* **336**, 1–166 (2000).
10. A. Slobodeniuk, I. Levkivskyi, E. Sukhorukov, *Phys. Rev. B Condens. Matter Mater. Phys.* **88**, 165307 (2013).
11. K. A. Matveev, *Phys. Rev. B Condens. Matter* **51**, 1743–1751 (1995).
12. Y. Nazarov, *Phys. Rev. Lett.* **82**, 1245–1248 (1999).
13. S. Jezouin *et al.*, *Nature* **536**, 58–62 (2016).
14. S.-Y. Lee, H.-W. Lee, H.-S. Sim, *Phys. Rev. B Condens. Matter Mater. Phys.* **86**, 235444 (2012).
15. E. Sivre *et al.*, *Nat. Phys.* **14**, 145–148 (2018).
16. A. Whiticar *et al.*, arXiv:1902.07085 [cond-mat.mes-hall] (2019).
17. Y. Ji *et al.*, *Nature* **422**, 415–418 (2003).
18. P. Roulleau *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **76**, 161309 (2007).
19. L. Litvin, H. Tranitz, W. Wegscheider, C. Strunk, *Phys. Rev. B Condens. Matter Mater. Phys.* **75**, 033315 (2007).
20. E. Bieri *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **79**, 245324 (2009).
21. H. Duprez *et al.*, *Phys. Rev. X* **9**, 021030 (2019).
22. I. Neder, M. Heiblum, Y. Levinson, D. Mahalu, V. Umansky, *Phys. Rev. Lett.* **96**, 016804 (2006).
23. N. Ofek *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5276–5281 (2010).
24. C. Bäuerle *et al.*, *Phys. Rev. Lett.* **95**, 266805 (2005).
25. Materials and methods are available as supplementary materials.
26. Z. Iftikhar *et al.*, *Nat. Commun.* **7**, 12908 (2016).
27. P. Roulleau *et al.*, *Phys. Rev. Lett.* **102**, 236802 (2009).
28. C. Altimiras *et al.*, *Phys. Rev. Lett.* **105**, 226804 (2010).
29. I. Sivan *et al.*, *Nat. Commun.* **7**, 12184 (2016).
30. B. I. Halperin, A. Stern, I. Neder, B. Rosenow, *Phys. Rev. B Condens. Matter Mater. Phys.* **83**, 155440 (2011).
31. R. Schuster *et al.*, *Nature* **385**, 417–420 (1997).
32. A. van Oudenaarden, M. Devoret, Y. Nazarov, J. Mooij, *Nature* **391**, 768–770 (1998).
33. M. de Jong, C. Beenakker, *Physica A* **230**, 219–248 (1996).
34. A. Bertoni, P. Bordone, R. Brunetti, C. Jacoboni, S. Reggiani, *Phys. Rev. Lett.* **84**, 5912–5915 (2000).
35. R. Ionicioiu, G. Amaratunga, F. Udrea, *Int. J. Mod. Phys. B* **15**, 125 (2001).
36. T. M. Stace, C. H. Barnes, G. J. Milburn, *Phys. Rev. Lett.* **93**, 126804 (2004).
37. E. Bocquillon *et al.*, *Ann. Phys. (Berlin)* **526**, 1–30 (2014).
38. D. Glattli, P. Roulleau, *Phys. Status Solidi, B Basic Res.* **254**, 1600650 (2017).
39. C. Bäuerle *et al.*, *Rep. Prog. Phys.* **81**, 056503 (2018).
40. H. Duprez, E. Sivre, A. Anthore, A. Aassime, A. Cavana, U. Gennser, F. Pierre, Data displayed in “Transmitting the quantum state of electrons across a metallic island with Coulomb interaction”, Zenodo (2019); doi 10.5281/zenodo.3528156.

ACKNOWLEDGMENTS

We thank P. Brouwer, L. Glazman, C. Mora, H. Sim, and E. Sukhorukov for illuminating discussions. **Funding:** This work was supported by the French RENATECH network, the national French program ‘Investissements d’Avenir’ (Labex NanoSaclay, ANR-10-LABX-0035), and the French National Research Agency (projects QuTherm, ANR-16-CE30-0010, and SIM-CIRCUIT, ANR-18-CE47-0014-01). **Author contributions:** E.S. and H.D. performed the experiment and analyzed the data with help from A.Aa., A.An., and F.P.; F.P. fabricated the sample with assistance from E.S and H.D.; A.C. and U.G. grew the 2DEG; F.P. led the project and wrote the manuscript with input from A.Aa., A.An., E.S., H.D., and U.G **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The data shown in the paper are available at Zenodo (*40*). Correspondence and requests for materials should be addressed to F.P. (frederic.pierre@c2n.upsaclay.fr).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1243/suppl/DC1
Materials and Methods
Figs. S1 to S5

24 January 2019; accepted 6 November 2019
10.1126/science.aaw7856

CANCER

Embryonal precursors of Wilms tumor

Tim H. H. Coorens[1], Taryn D. Treger[1,2,3], Reem Al-Saadi[4,5], Luiza Moore[1,2], Maxine G. B. Tran[6,7], Thomas J. Mitchell[1,2,8], Suzanne Tugnait[4], Christine Thevanesan[4], Matthew D. Young[1], Thomas R. W. Oliver[1,2,9], Minou Oostveen[4,5], Grace Collord[1,2,3], Patrick S. Tarpey[2], Alex Cagan[1], Yvette Hooks[1], Mark Brougham[10], Ben C. Reynolds[11], Giuseppe Barone[5], John Anderson[4,5], Mette Jorgensen[5], G. A. Amos Burke[2,3], Johannes Visser[2], James C. Nicholson[2,3], Naima Smeulders[5], Imran Mushtaq[5], Grant D. Stewart[2,8], Peter J. Campbell[1], David C. Wedge[12,13], Iñigo Martincorena[1], Dyanne Rampling[5], Liz Hook[2,9], Anne Y. Warren[2,9], Nicholas Coleman[2,9], Tanzina Chowdhury[5], Neil Sebire[4,5], Jarno Drost[14], Kourosh Saeb-Parsy[2,8], Michael R. Stratton[1], Karin Straathof[4,5], Kathy Pritchard-Jones[4,5], Sam Behjati[1,2,3]*

Adult cancers often arise from premalignant clonal expansions. Whether the same is true of childhood tumors has been unclear. To investigate whether Wilms tumor (nephroblastoma; a childhood kidney cancer) develops from a premalignant background, we examined the phylogenetic relationship between tumors and corresponding normal tissues. In 14 of 23 cases studied (61%), we found premalignant clonal expansions in morphologically normal kidney tissues that preceded tumor development. These clonal expansions were defined by somatic mutations shared between tumor and normal tissues but absent from blood cells. We also found hypermethylation of the *H19* locus, a known driver of Wilms tumor development, in 58% of the expansions. Phylogenetic analyses of bilateral tumors indicated that clonal expansions can evolve before the divergence of left and right kidney primordia. These findings reveal embryonal precursors from which unilateral and multifocal cancers develop.

Adult cancers typically arise as a consequence of aging and mutagen exposure, at times through the generation of precancerous clonal expansions. Examples of these precancerous states include Barrett's esophagus, clonal hematopoiesis, and colonic polyps. It is unknown whether childhood tumors, which are thought to result from aberrant fetal development, likewise arise from precancerous clonal expansions (Fig. 1A). To address this question, we studied Wilms tumor (nephroblastoma), the most common kidney cancer of childhood. Wilms tumor is a prototypical embryonal malignancy of infants and young children (*1*). It arises from abnormal fetal nephrogenesis, which it resembles morphologically (*1*) and transcriptionally (*2*). It occurs sporadically or in the context of bilateral tumors, multifocal lesions, urogenital developmental disorders, or overgrowth syndromes (*1*).

[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK. [2]Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK. [3]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. [4]UCL Great Ormond Street Institute of Child Health, London WC1N 1EH, UK. [5]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK. [6]UCL Division of Surgery and Interventional Science, Royal Free Hospital, London NW3 2PS, UK. [7]Specialist Centre for Kidney Cancer, Royal Free Hospital, London NW3 2PS, UK. [8]Department of Surgery, University of Cambridge, Cambridge CB2 0QQ, UK. [9]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. [10]Department of Haematology and Oncology, Royal Hospital for Sick Children, Edinburgh EH9 1LF, UK. [11]Department of Paediatric Nephrology, Royal Hospital for Children, Glasgow G51 4TF, UK. [12]Big Data Institute, University of Oxford, Oxford OX3 7LF, UK. [13]Oxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford OX3 9DU, UK. [14]Princess Máxima Center for Pediatric Oncology, Oncode Institute, 3584 CS Utrecht, Netherlands.
***Corresponding author. Email: sb31@sanger.ac.uk**

To identify potential precursors of Wilms tumor, we used somatic mutations to infer the phylogenetic relationship between cancers and corresponding normal tissues (kidney and blood). We analyzed 229 whole-genome sequences of 54 individuals: 23 children with Wilms tumor, 16 parents of affected children, three children with other types of kidney cancer (congenital mesoblastic nephroma, malignant rhabdoid tumor), 10 adults with clear cell renal cell carcinoma (ccRCC), and two adults without renal tumors (one kidney transplant patient and one kidney obtained at autopsy; table S1). We called base substitutions against the reference human genome and extracted mosaic mutations from each set of donor-related tissues. We validated the method for calling mosaic mutations by sequencing parental germline DNA, resequencing tissues, and inspecting raw data (fig. S1). On the basis of the variant allele frequency (VAF) and distribution of mutations across related tissues, we built phylogenetic trees of tumor development. We supplemented DNA data with analyses of RNA sequences and genome-wide methylation patterns (table S1).

Our discovery cohort consisted of three children with unilateral Wilms tumor. We sampled tumors, blood, and histologically normal kidney tissues from the same individuals (table S1 and fig. S2). As expected, whole-genome sequences revealed mosaic mutations attributable to the first cell divisions of the fertilized egg (fig. S3 and table S2) (*3*, *4*). In two cases, we also detected mosaic mutations in normal kidneys that were present in the corresponding cancer but absent from blood (Fig. 1, B to D, and table S2), indicating that the tumors had arisen from that particular normal kidney tissue.

Several features of these mutations showed that they defined clonal expansions in normal kidney tissue, as illustrated by patient PD37272 (Fig. 1, B to D). The VAFs of mutations in the normal tissue of this kidney, variants 3 to 5 (Fig. 1C), were as high as 44%, which suggested that the mutation was present in 88% of all cells in the biopsy. Mutations 3 to 5 were present in the two parenchymal biopsies (i.e., cortex and medulla) but were absent from blood cell DNA, deeply sequenced to 106× genome-wide (fig. S4). Similarly, mutations 3 to 5 were undetectable in renal pelvis, which is embryologically derived from a different lineage than kidney parenchyma (*5*). Furthermore, the VAF of early embryonic mutations, variants 1 and 2, was inflated in parenchyma and in tumors (Fig. 1D). Such inflation of early embryonic mutations is a feature of tissues that contain a clonal expansion of a single cell (fig. S3). By contrast, in tissues devoid of a major clone, such as renal pelvis and blood, the VAFs of early embryonic mutations were not inflated (Fig. 1D). Thus, these normal tissue variants 3 to 5 demonstrate the presence of clonal expansions within kidney parenchyma, which we termed clonal nephrogenesis, accounting for up to 88% of cells sampled in the cortex.

To further study and validate our discovery of clonal nephrogenesis as an antecedent of Wilms tumor, we studied another 20 cases: 15 unilateral tumors with normal tissue biopsies curated through a British childhood renal tumor study (IMPORT), four cases of bilateral Wilms tumor, and one tumor with 10 normal tissue biopsies (table S1 and fig. S5). Within the entire group of 23 children (discovery and validation cohorts), we found evidence of clonal nephrogenesis in 10 of 19 children with unilateral disease (53%) and in all four children with bilateral cancers (Fig. 2, A and B). The presence of clonal nephrogenesis was further substantiated by the significant ($P < 0.01$, Wilcoxon signed-rank test) inflation of VAFs of early embryonic variants (fig. S6). There were no copy number changes detected in normal tissues by three different methods.

Conceivably these findings could be due to tumor infiltration into normal tissue not visible histologically (fig. S2) or to cross-contamination of DNAs. This explanation is implausible, as contamination would manifest as shared variants at a low VAF, rather than select mutations at a high VAF. We statistically excluded the possible contribution of tumor infiltration and contamination by using a binomial mixture model on the observed base counts of tumor mutations in the normal samples (fig. S7).

Next, we investigated whether clonal nephrogenesis represents the normal clonal architecture of human nephrons by three approaches. First, using laser capture microscopy (LCM),

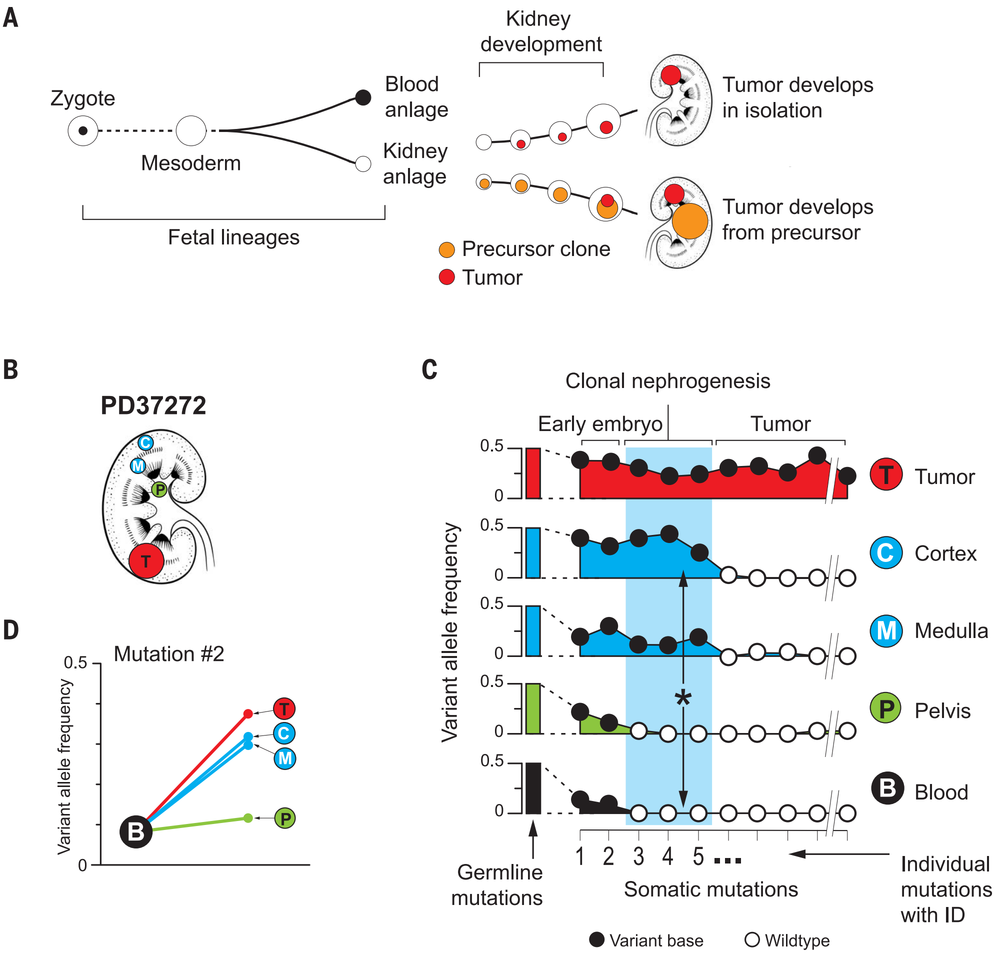


Fig. 1. Precursor clonal expansions in normal human kidneys. (**A**) Wilms tumor can arise during embryogenesis on the background of an otherwise normal kidney (the traditional view) or from precursor lesions residing in normal tissues, as found here. (**B**) Overview of tissue sampling in the kidney of patient PD37272. (**C**) Somatic mutations can be timed using their VAF across the corresponding normal tissues, which is higher for earlier mutations. If the mutation is present in tumor, kidney, and blood, it is an early embryonic mutation (mutations 1 and 2). If it is present in kidney samples and tumor only, it is clonal nephrogenic (mutations 3 to 5, marked by asterisk). If it is only in the tumor, it is labeled as such. White and black circles indicate whether the observed VAF is insignificant (white) or significant (black), $P < 0.001$ (test of presence using beta-binomial overdispersion). (**D**) The VAF for the last embryonic mutation in kidney samples and tumor compared with blood.

we excised glomeruli ($n = 7$) and proximal and distal tubules ($n = 15$) from the kidneys of three individuals (Fig. 2, C and D). We subjected these LCM cuts to whole-genome sequencing using an established method for generating low-input DNA libraries (*6*, *7*). Kidney tissues were obtained at autopsy (one case) or from normal portions of kidneys affected elsewhere by ccRCC (two cases). Our analysis of somatic mutations across these LCM cuts revealed a VAF distribution (Fig. 2E) that is inconsistent with the monoclonal organization seen, for example, in endometrial glands or colonic crypts (*6*, *7*). Second, we assessed whether mutations were commonly shared between renal tumors and surrounding normal kidney tissue. We studied childhood congenital mesoblastic nephroma (two tumors and six normal kidney samples), childhood malignant rhabdoid tumor (one cancer and one normal kidney sample), and adult ccRCC (eight cancers, including one bilateral case, and 15 normal tissues). Applying the same analysis pipeline, we identified early embryonic mutations shared between tumor, normal kidney tissues, and blood (fig. S8). However, we did not find mutations shared only between tumor and normal tissue (Fig. 2B); this finding shows that such mutations were specific to, and enriched in, Wilms tumor ($P < 0.001$, Fisher exact test). Of particular relevance were renal cell carcinomas, which, like Wilms tumor, are derived from nephrons. If normal embryological clonal dynamics typically had generated clonal expansions, we would have expected to find clonal nephrogenesis in ccRCC cases. Third, we examined all developmental mutations of normal kidney tissues listed thus far, supplemented by an additional 18 biopsies obtained from bilateral kidneys that had been declined for transplantation (Fig. 2B). We analyzed somatic mutations present in kidney tissue and absent from nonrenal tissue, irrespective of whether they were shared with tumors. Collectively, these analyses of 77 normal kidney biopsies revealed that variants of tissues without clonal nephrogenesis have a significantly lower VAF distribution than clonal nephrogenesis mutations ($P < 0.001$, Wilcoxon rank-sum test; Fig. 2B). Taken together, these results indicate that clonal nephrogenesis represents aberrant kidney development.

A central question raised by our findings is whether cancer-causing (driver) events underpin clonal nephrogenesis. Notably, the clonal nephrogenesis mutations ($n = 66$) (table S2 and fig. S8) were noncoding (64 of 66) or did not generate plausible oncogenic events. We searched further for driver events among germline and somatic DNA mutations, in transcriptomes, and, where available, in methylation patterns. We found hypermethylation of the *H19* locus in 7 of 12 (58%) normal kidney tissues with clonal nephrogenesis. In contrast, we did not detect it in normal kidney tissues without clonal nephrogenesis or in blood (Fig. 2, A and F), except for the blood of a child (PD40738) with Beckwith-Wiedemann syndrome (fig. S9). *H19* hypermethylation is an established driver event in the pathogenesis of Wilms tumor and is thought to operate by disrupting the epigenetic regulation of growth-promoting genes that reside on chromosome 11p15 (*8*–*10*). The degree of hypermethylation of *H19* correlated with the VAF of clonal nephrogenesis, indicating that hypermethylation pervaded the entire clone (Fig. 2G). In the five wild-type samples, hypermethylation of *H19* may therefore have evaded detection because of a small clone size (Fig. 2A). We also studied the KvDMR1 locus on chromosome 11p15.5 and determined that its methylation state was unchanged in clonal nephrogenesis. Hypomethylation of KvDMR1 underlies overgrowth

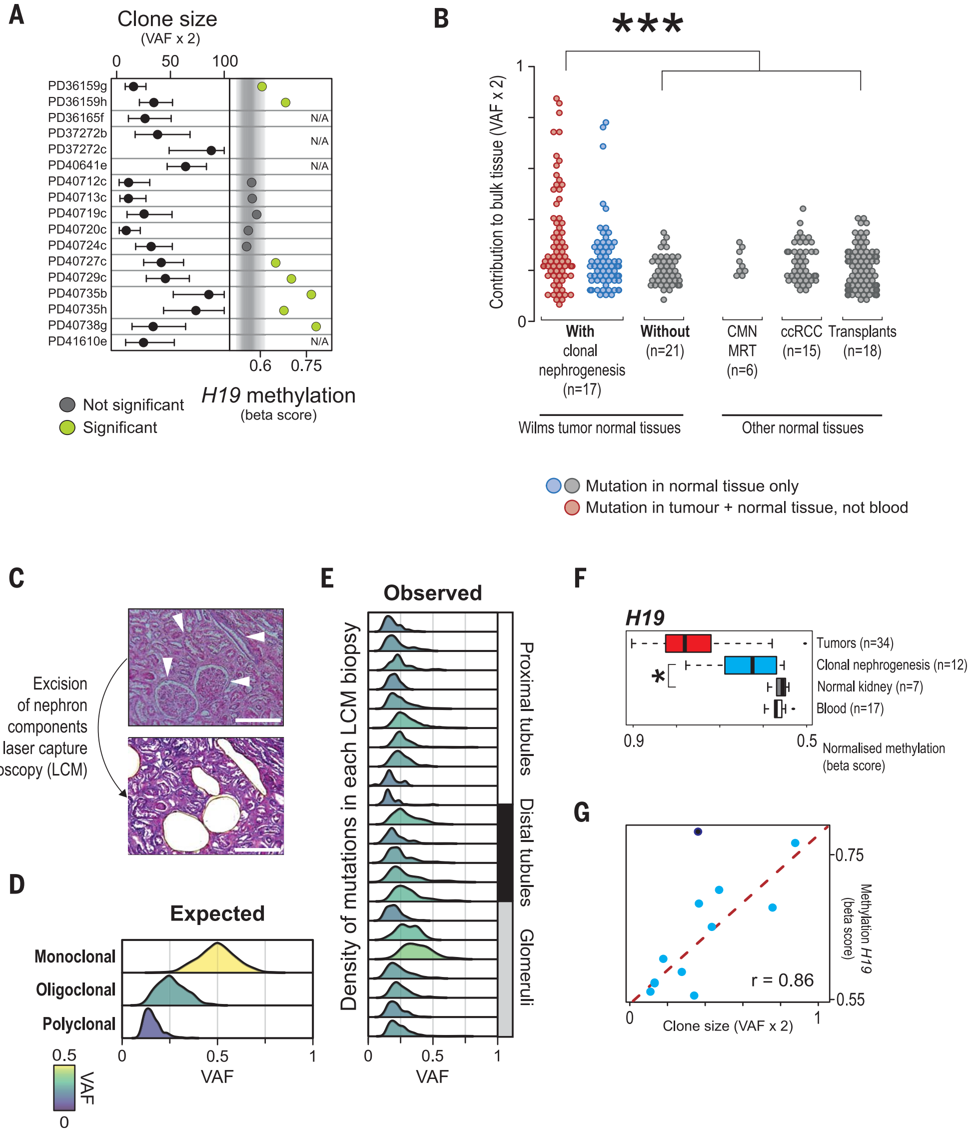


Fig. 2. Clonal nephrogenesis and H19 hypermethylation. (**A**) Sizes of nephrogenic clones, as predicted by twice the VAF of the most prominent nephrogenic mutation, alongside corresponding methylation values of the *H19* locus. Green symbols indicate significant deviation ($P < 0.05$, Wilcoxon rank-sum test) from the background methylation distribution (gray shaded area) as obtained from normal kidney samples without clonal nephrogenesis. N/A, no methylation data available. (**B**) Mutations present in samples obtained from normal kidneys but absent in matched blood. Only in Wilms tumor were some of these mutations shared with the corresponding tumor. In the presence of clonal nephrogenesis, the VAF distribution of these mutations was significantly elevated (***$P < 0.001$, Wilcoxon rank-sum test). (**C**) Histology images showing components (arrowheads) of the human nephron excised by laser capture microscopy. Scale bar, 250 μm. (**D**) VAF simulations to derive expected distributions depending on clonality of a tissue: monoclonal origin (~*Bin*(*n*~*Pois*(*cov*), *P* = 0.5)), oligoclonal origin (~*Bin*(*n*~*Pois*(*cov*), *P* = 0.3)), or polyclonal origin (~*Bin*(*n*~*Pois*(*cov*), *P* = 0.1)), where *cov* = coverage = 40×. (**E**) VAF distributions for 22 microdissected samples (10 proximal tubules, five distal tubules, seven glomeruli) from three patients, one rapid autopsy donor, and two ccRCC patients. Color indicates the underlying maximum likelihood VAF as predicted by a truncated binomial mixture model. (**F**) Group-level methylation beta values of *H19* (*$P < 0.05$, Wilcoxon rank-sum test). (**G**) Relationship between predicted clone sizes from nephrogenic mutation (see Fig. 2A) and the methylation level of *H19*. The black dot represents patient PD40738g, who is affected by germline *H19* hypermethylation (omitted from correlation and linear regression).

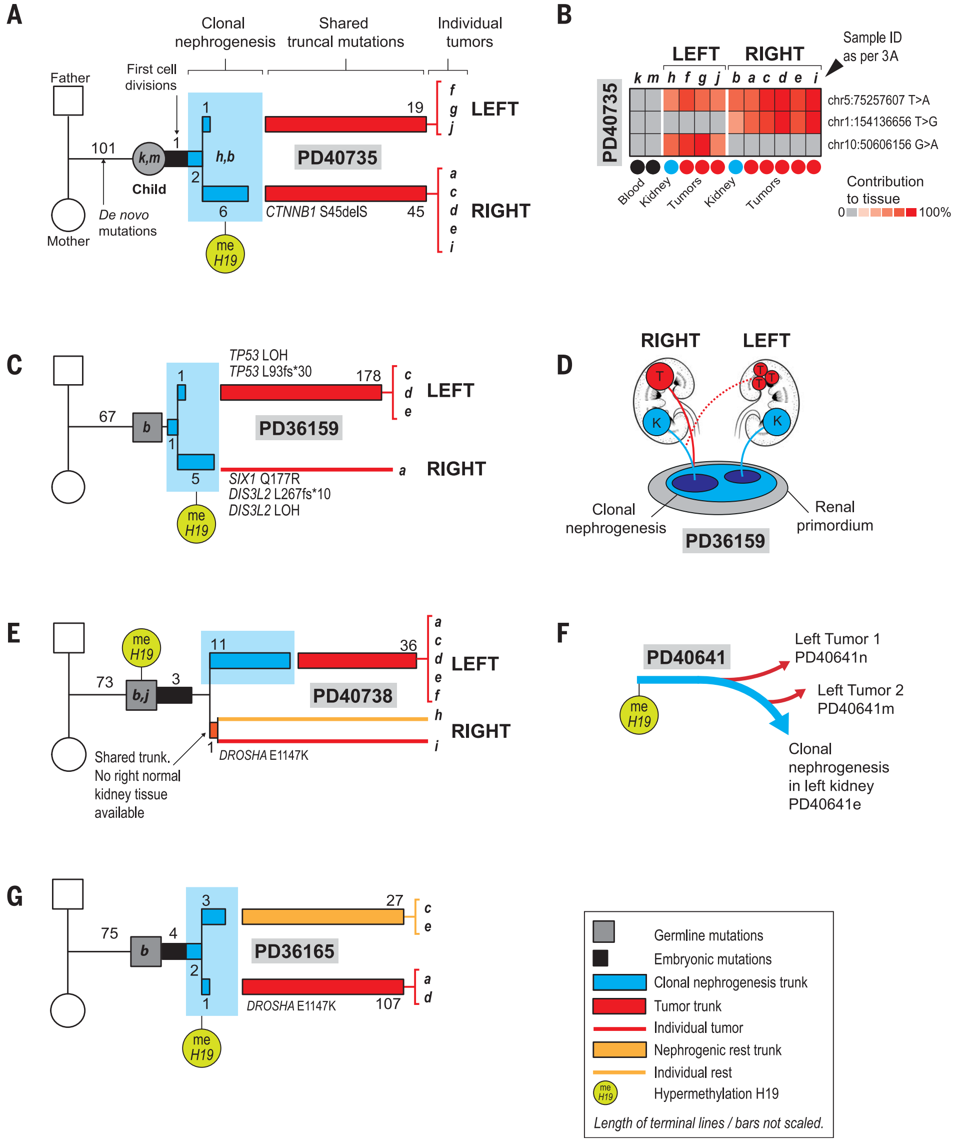
syndromes that confer only a minimal predisposition to Wilms tumor (*11*). We did not find any further driver events accounting for clonal nephrogenesis, despite re-interrogation by exome sequencing of 15 of 17 tissues with clonal nephrogenesis. Gene expression profiles, including fetal transcripts, did not differ between normal renal tissues that displayed or did not display clonal nephrogenesis (fig. S9). Similarly, global methylation patterns did not differ between these two groups (fig. S9).

The timing of the emergence of clonal nephrogenesis during development could be defined in three children from whom we obtained bilateral tumors. In patients PD40735 and PD36159, left and right tumors were derived from the same trunk of clonal nephrogenesis (Fig. 3, A to D). In the third patient, PD40378, all five left tumors, but not the right neoplasms, were related to clonal nephrogenesis on the left (Fig. 3E). These findings indicate that in two children with bilateral cancers, clonal nephrogenesis must have arisen before left and right kidney primordia diverged, early in embryogenesis (*5*). For unilateral tumors, the timing of the occurrence of clonal nephrogenesis remains unclear. It may have evolved before the kidney was formed or thereafter, followed by a "clonal sweep" of clonal nephrogenesis replacing kidney tissue.

In five cases, we sampled multiple neoplasms of the same kidney, which revealed two configurations of tumor development (Fig. 3). Tumors either were derived from a shared trunk that had emerged from clonal nephrogenesis (Fig. 3, A to E) or arose independently and successively from clonal nephrogenesis, which suggests a sustained potential for the latter to spawn tumors (Fig. 3, F and G). For example, patient PD36165 presented with two lesions at opposing poles of the kidney: a nephrogenic rest and a Wilms tumor. Cancer and the nephrogenic rest had emerged from the same ancestral clone at different time points, followed by clonal diversification within each (Fig. 3G).

Finally, we compared the genetic alterations in Wilms tumor with and without clonal nephrogenesis. This revealed a near–mutual exclusivity of loss of heterozygosity (LOH) of chromosome 11p15 and clonal nephrogenesis. In tumors with clonal nephrogenesis, LOH of 11p15 was mostly absent ($P = 0.009$, Fisher exact test) (fig. S10). This suggests that there may be two distinct pathways for the pathogenesis of Wilms tumor that both involve dysregulation of 11p15 genes as a driver (Fig.

Fig. 3. Phylogenies of bilateral and multifocal Wilms tumor. (**A**, **C**, **E**, and **G**) For each tumor, the phylogeny of shared mutations is shown including de novo germline mutations, embryonic mutations, mutations demarcating clonal nephrogenesis, and tumor mutations. Numbers refer to the number of substitutions defining each developmental trunk. Truncal driver events are detailed. (**B**) Heat map showing the contribution of a mutation to the sample PD40735 shown in (A). The pattern of shared mutations reveals a split between left and right kidney, which is obeyed by both tumor and normal samples. (**D**) As revealed by the pattern of shared mutations, in patient PD36159 the left tumor is more closely related to the right branch of clonal nephrogenesis than to the left branch. (**F**) Two mutations indicate the independent emergence of tumors at different time points from the nephrogenic clone in patient PD40641. As shown in (G), tumor and nephrogenic rest in patient PD36165 both originated from clonal nephrogenesis despite being situated at opposing kidney poles.



1A). Accordingly, cancers may arise either directly in isolation through LOH of 11p15, or indirectly via clonal nephrogenesis with perturbation of 11p15 by hypermethylation of *H19*.

Our findings show that clonal expansions in histologically normal tissue are an atypical outcome of renal development that commonly antedates Wilms tumor. We demonstrated a direct phylogenetic link between clonal expansions, *H19* hypermethylation, and the formation of cancer, thus identifying clonal nephrogenesis as an epigenetic progenitor of cancer comprising "neoplasia-ready cells" (*12*). In contrast to precursors of adult cancer, clonal nephrogenesis generated histologically and functionally normal tissues, which in the most pronounced cases occupied the bulk of renal tissues. If future work shows that the extent of clonal nephrogenesis is a marker of malignant potential and risk of cancer recurrence, this information potentially could be used to guide treatment and surveillance of patients with Wilms tumor. Moreover, if we were able to manipulate the neoplastic potential of clonal nephrogenesis, prevention of Wilms tumor could become feasible. Collectively, our findings portray Wilms tumor as an insurrection on the background of a premalignant tissue bed, rather than a clearly demarcated neoplasm in an otherwise normal polyclonal kidney. We speculate that embryonal clonal expansions, perhaps also driven by epigenetic mechanisms, may be a common phenomenon in childhood cancer.

REFERENCES AND NOTES

1. T. D. Treger, T. Chowdhury, K. Pritchard-Jones, S. Behjati, *Nat. Rev. Nephrol.* **15**, 240–251 (2019).
2. M. D. Young *et al.*, *Science* **361**, 594–599 (2018).
3. S. Behjati *et al.*, *Nature* **513**, 422–425 (2014).
4. Y. S. Ju *et al.*, *Nature* **543**, 714–718 (2017).
5. K. M. Short, I. M. Smyth, *Nat. Rev. Nephrol.* **12**, 754–767 (2016).
6. H. Lee-Six *et al.*, *Nature* **574**, 532–537 (2019).
7. L. Moore *et al.*, bioRxiv 505685 [preprint]. 24 December 2018.
8. T. Moulton *et al.*, *Nat. Genet.* **7**, 440–447 (1994).
9. K. Okamoto, I. M. Morison, T. Taniguchi, A. E. Reeve, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5367–5371 (1997).
10. J. Charlton *et al.*, *Genome Med.* **7**, 11 (2015).
11. W. N. Cooper *et al.*, *Eur. J. Hum. Genet.* **13**, 1025–1032 (2005).
12. A. P. Feinberg, R. Ohlsson, S. Henikoff, *Nat. Rev. Genet.* **7**, 21–33 (2006).

ACKNOWLEDGMENTS

We thank the research nursing staff at Cambridge University Hospitals, the Royal Hospital for Sick Children (Edinburgh), and the Royal Hospital for Children (Glasgow) as well as all IMPORT investigators. We thank M. Gerstung and A. J. Lawson for critical review of the manuscript. We are indebted to our little and older patients and their families for participating in our research. **Funding:** This project was principally funded by the Little Princess Trust, the St. Baldrick's Foundation (Robert J. Arceci International Award to S.B.), and Wellcome (fellowship to S.B.; Sanger core funding). Additional funding was received from CRUK (IMPORT study; fellowship to T.J.M.; Cambridge Centre), NIHR (Biomedical Research Centre Great Ormond Street; Cambridge Human Research Tissue Bank; Oxford Biomedical Research Centre; fellowship to T.R.W.O.), the Royal College of Surgeons of England (fellowship to T.J.M.), Wellcome (fellowship to T.H.H.C. and K.S.), Great Ormond Street Hospital Children's Charity

(R.A.-S., K.P.-J.), and Li Ka Shing foundation (D.C.W.). J.D. acknowledges funding from the Alpe d'HuZes foundation/KWF Dutch Cancer Society Bas Mulder Award (#10218) and the Oncode Institute. **Author contributions:** S.B. conceived of the experiment. T.H.H.C., T.D.T., and S.B. analyzed data. Statistical expertise was provided by M.D.Y., D.C.W., and I.Ma. M.D.Y., T.J.M., G.C., P.S.T., P.J.C., M.R.S., I.Ma., and K.P.-J. contributed to discussion. L.M. and M.G.B.T. performed laser capture microscopy experiments. Samples were curated and/or experiments were performed by R.A.-S., C.T., S.T., M.O., Y.H., M.B., B.C.R., G.B., J.A., M.J., G.A.A.B., J.V., J.C.N., N.S., K.S.-P., G.D.S., K.S., T.C., I.Mu., J.D., and K.P.-J. Pathological expertise was provided by L.H., T.R.W.O., D.R., A.Y.W., N.C., and N.S. A.C. created kidney illustrations. T.H.H.C., T.D.T., and S.B. wrote the manuscript. S.B. directed the study. **Competing interests:** G.D.S. is a paid consultant for Pfizer, Merck, EUSA Pharma, and Cambridge Medical Robotics. All other authors declare no competing interests. **Data and materials availability:** Raw sequencing data have been deposited in the European Genome-phenome Archive (EGA) under study ID EGAD00001004774. The bespoke code used for the analyses is available on GitHub (https://github.com/TimCoorens/ClonalNephrogenesis).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1247/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S8
References (*13*–*30*)

View/request a protocol for this paper from *Bio-protocol*.

25 February 2019; resubmitted 4 July 2019
Accepted 6 November 2019
10.1126/science.aax1323

HUMAN RETINA

Functional diversity of human intrinsically photosensitive retinal ganglion cells

Ludovic S. Mure[1]*, Frans Vinberg[2], Anne Hanneken[3], Satchidananda Panda[1]*

Intrinsically photosensitive retinal ganglion cells (ipRGCs) are a subset of cells that participate in image-forming and non–image-forming visual responses. Although both functional and morphological subtypes of ipRGCs have been described in rodents, parallel functional subtypes have not been identified in primate or human retinas. In this study, we used a human organ donor preparation method to measure human ipRGCs' photoresponses. We discovered three functional ipRGC subtypes with distinct sensitivities and responses to light. The response of one ipRGC subtype appeared to depend on exogenous chromophore supply, and this response is conserved in both human and mouse retinas. Rods and cones also provided input to ipRGCs; however, each subtype integrated outer retina light signals in a distinct fashion.

In mammals, intrinsically photosensitive retinal ganglion cells (ipRGCs) are a subset of retinal ganglion cells that express the photopigment melanopsin, which renders them intrinsically photosensitive. These cells participate in a set of light responses that include circadian entrainment, pupillary light reflex (PLR), and the modulation of sleep or alertness and mood as well as some aspects of vision (*1*). In rodents, six different morphological and at least three functional ipRGC types have been described (*2*–*4*). In addition to intrinsic photosensitivity, the ipRGCs also mediate rod- and cone-initiated photoresponses, which expand the range of sensitivity for ipRGCs to dim light. Melanopsin is also found in the human retina (*5*), where it is responsible for the suppression of nocturnal melatonin and PLR (table S1). However, there is no report of previous direct assessment of ipRGC function in humans.

To assess the intrinsic photoresponses of human RGCs, we performed extracellular electrophysiological recordings of freshly harvested human retinas from five donors (three females, two males, 57.2 ± 8.8 years old; table S2). Small pieces of retina were placed on a multielectrode array, and the recording medium was supplemented with synaptic blockers to block excitatory rod and cone input to RGCs. In response to a 30-s pulse of monochromatic blue light (470 nm), photoresponsive cells were found in retinas from each of the five donors (Fig. 1A). Intrinsic responses to light were slow, sustained over the 30-s stimulation, and outlasted the stimulus by several seconds after the light was switched off (Fig. 1, A and B, and fig. S1A).

For each retina sample, we identified photoresponsive cells at an average density of 2.47 cells/mm^2 (fig. S1B), close to the lower range of the reported density of melanopsin-immunopositive RGCs in human retinas (from ~3 to 40 cells/mm^2) (*6*–*9*). Furthermore, intrinsic photoresponses from the retinas were reversibly inhibited by opsinamide (AA92593), a specific inhibitor of melanopsin (*10*) (Fig. 1, C

[1]Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]John A. Moran Eye Center, University of Utah, 65 Mario Capecchi Drive (S3140), Salt Lake City, UT 84132, USA. [3]Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.
***Corresponding author. Email: lmure@salk.edu (L.S.M.); panda@salk.edu (S.P.)**

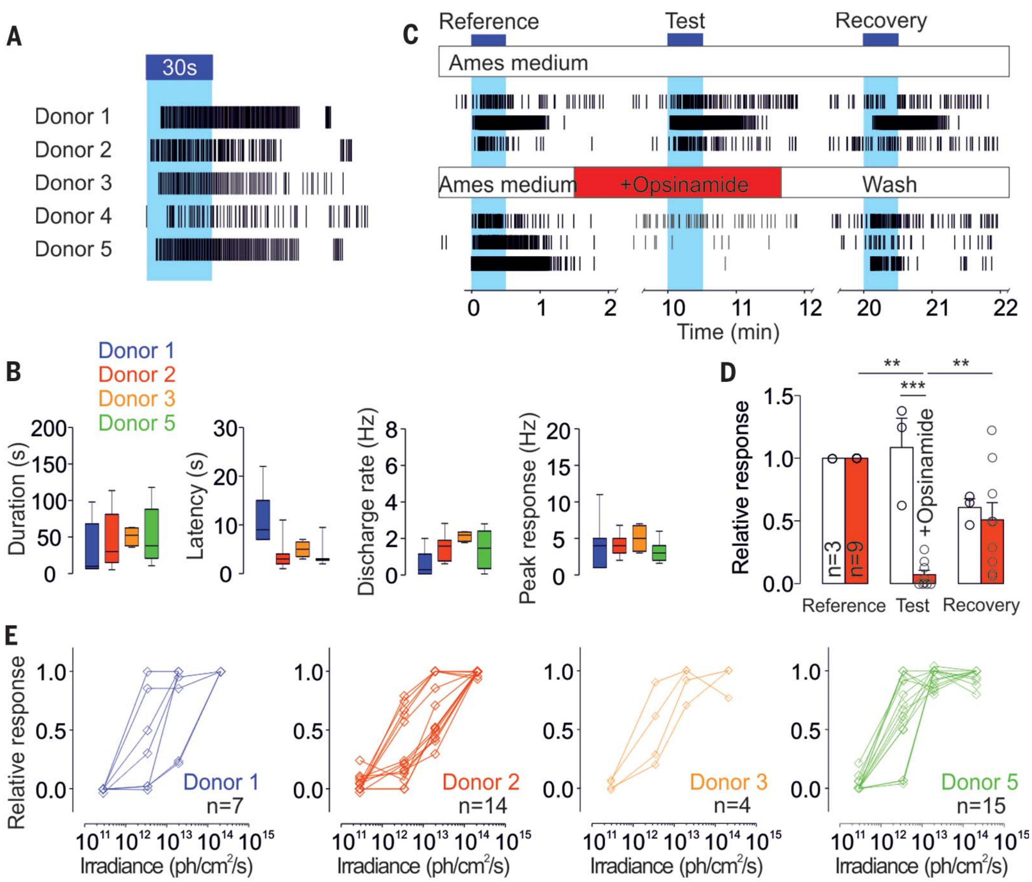


Fig. 1. A subset of human RGCs is intrinsically photosensitive. (**A**) Representative individual ipRGC spike trains to a 30-s light pulse (~10^{13} photons/cm^2 per second, 470 nm) from five different human donors. (**B**) Average ipRGCs' response latency, duration, and amplitude in each donor (*n* = 7, 15, 4, and 15 for donors 1, 2, 3, and 5, respectively, where *n* is the number of cells). Representative individual ipRGCs spike trains (**C**) and average responses (**D**) to three identical 30-s light pulses (~10^{13} photons/cm^2 per second, 470 nm) from a control recording [*n* = 3; upper panel in (C), white histogram in (D)] or a recording with opsinamide (*n* = 9; lower panel in (C), red histogram in (D)] during the second stimulation [***P < 0.001, **P < 0.01, ANOVA (analysis of variance), Bonferroni post hoc test, one donor]. (**E**) Responses of ipRGCs stimulated for 30 s at different irradiances (from 2 × 10^{11} to 2 × 10^{14} photons/cm^{2}per second, 470 nm) in each donor (*n* = 7, 14, 4, and 15 for donors 1, 2, 3, and 5, respectively). ph, photons. Blue bars and blue background in (A) and (C) indicate light pulses.

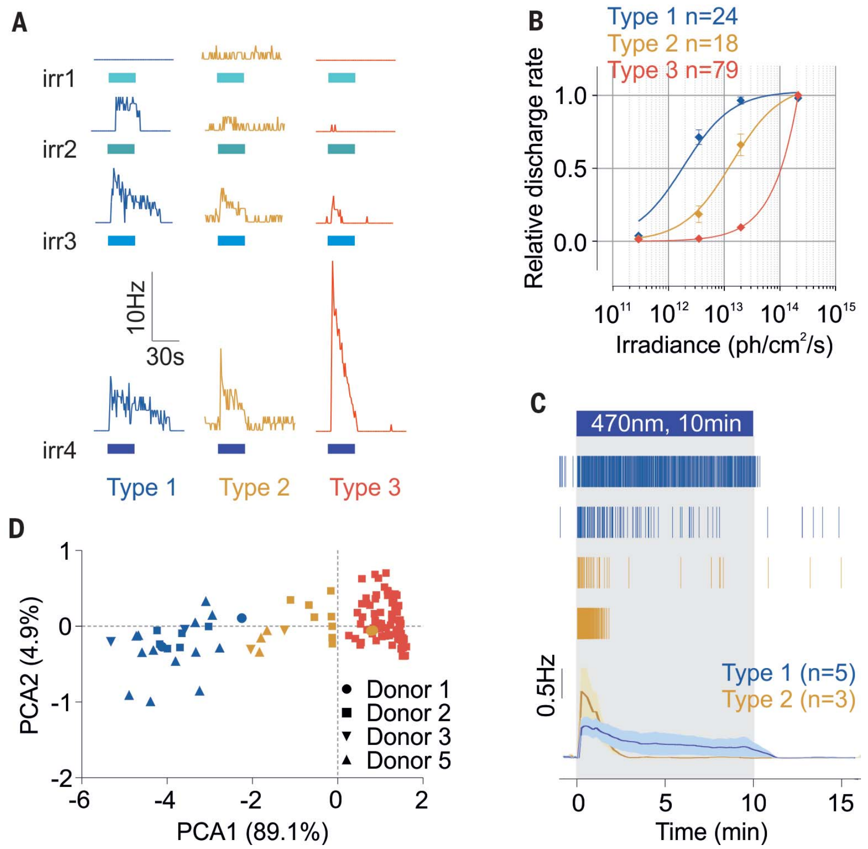


Fig. 2. Human ipRGCs display different subtypes. (**A**) Representative responses from type 1, 2, and 3 ipRGCs to increasing irradiance light pulses [30 s, 470 nm, irradiances (irr) 1, 2, 3, and 4: 2.9×10^{11}, 3.5×10^{12}, 2×10^{13}, and 2×10^{14} photons/cm^2 per second, respectively]. Blue bars indicate light pulses. (**B**) Corresponding dose–response curves (type 1, n = 24, four donors; type 2, n = 18, four donors; type 3, n = 79, one donor). Error bars indicate SEM. (**C**) Representative raster plots of type 1 and 2 ipRGCs and average traces in response to 10-min light stimulations (~2×10^{13} photons/cm^2 per second, 470 nm; type 1, n = 5, and type 2, n = 3). (**D**) Principal components of the ipRGCs' response parameters (sensitivity, latency, and duration; n = 121, four donors).

and D), supporting the notion that the intrinsic photosensitivity is mediated by melanopsin.

Next, we tested whether human ipRGCs, like their rodent counterparts, can sustain photoresponses under prolonged illumination (*11*, *12*). In response to a 10- or 20-min illumination, all ipRGCs responded for several minutes (fig. S1C), whereas 40 and 28% of ipRGCs sustained responses to the entire 10 or 20 min of light, respectively.

To assess human ipRGCs' sensitivity, we stimulated the retinas for 30 s at increasing irradiances and found similar profiles of sensitivity among donors (Fig. 1E). The ipRGC responses were rarely detectable at 2×10^{11} photons/cm^2 per second, and half-saturation sensitivities were recorded between 10^{12} and 10^{13} photons/cm^2 per second. Altogether, the response characteristics of human ipRGCs—low sensitivity, slow activation, sustained response during light stimulation, and delayed deactivation—were similar to those seen in the mouse (*13*), *Arvicanthis* (*14*), and nonhuman primate cells (*15*).

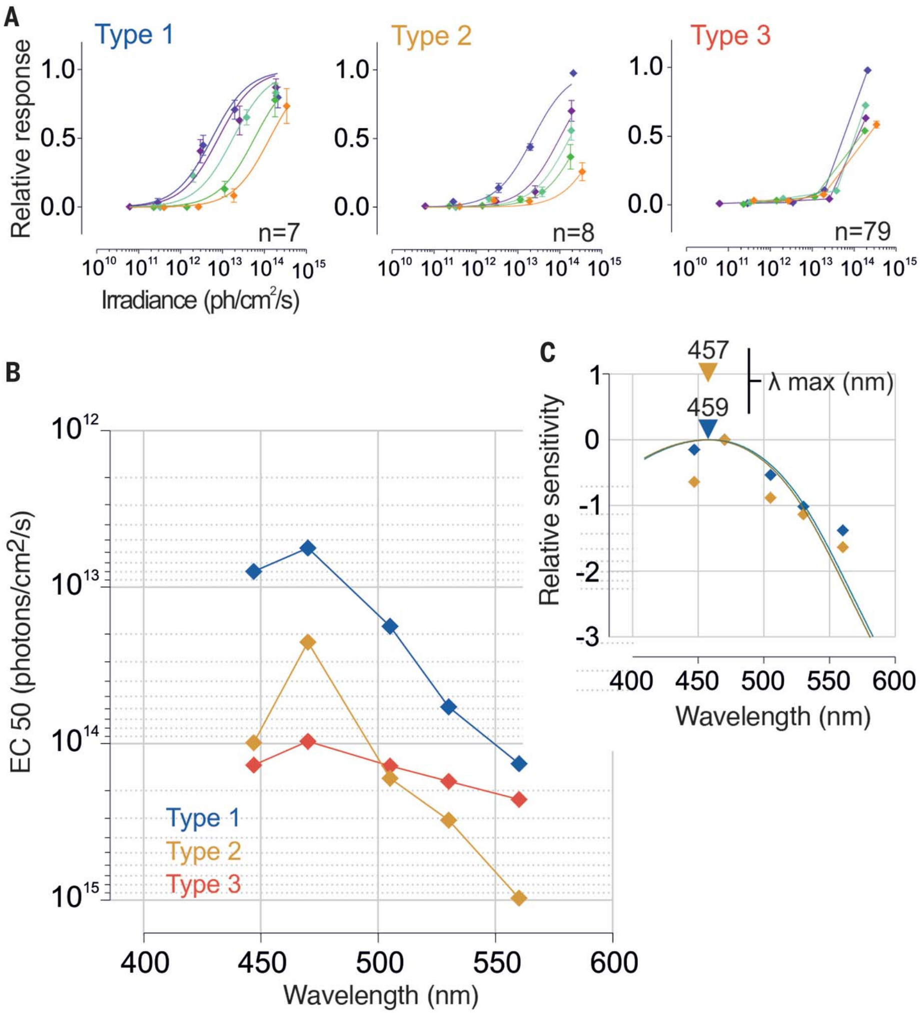
In mice, six subtypes of ipRGCs (M1 to M6) have been described on the basis of their morphologies, levels of expression of melanopsin, connectivity patterns, and photoresponses (*2*, *16*). Response sensitivity (Fig. 1E) seemed to delineate at least two different types of human ipRGCs. Principal components analysis (PCA) of response parameters (sensitivity, latency, and duration) showed that, independent of the donor, ipRGC responses tended to cluster into two groups, which we called type 1 and type 2 ipRGCs (Fig. 2A and fig. S2, A to C).

Type 1 ipRGCs were more sensitive to light, with a half-maximal response at ~2×10^{12} photons/cm^2 per second and a sustained light response that lasted 47 ± 10.9 s after a 30 s pulse of light (2.1×10^{14} photons/cm^2 per second) was turned off (Fig. 2B and fig. S2D). Type 2 ipRGCs were less sensitive, with a half-maximal response at ~1×10^{13} photons/cm^2 per second, and response termination 23.9 ± 10.7 s after lights were turned off (Fig. 2B and fig. S2D). At lower irradiance levels, type 2 ipRGCs exhibited a longer response latency to the test light pulse (fig. S2D). Overall, type 1 responses were recorded 50% more frequently than type 2 responses (1.6 ± 0.3 versus 1.1 ± 0.5 cell/mm^2, respectively) (fig. S2E). Altogether, the features of type 1 and type 2 ipRGCs suggest that they correspond to mouse ipRGC subtypes M1 and M2 (*1*) or type II and type III (*3*) (table S3). The type 1 ipRGCs also sustained longer responses under prolonged illumination, whereas type 2 cells were refractory after <5 min of illumination (Fig. 2C and fig. S2F).

Because some rodent ipRGCs may rely on the retinal pigment epithelium to supply the 11-*cis*-retinaldehyde (11-*cis*-retinal) melanopsin chromophore (*17*), we preincubated one retina sample in 11-*cis*-retinal for 1 hour before the recordings. In addition to type 1 and 2 responses, we also discovered a third type of ipRGC that clustered separately (Fig. 2, A and F, and fig. S2G). The cells exhibiting type 3 responses were distinct; they responded only to higher irradiances, but more strongly (Fig. 2, A and B, and fig. S2, C and H). Their response latencies were similar to those of types 1 and 2, but their response durations were shorter, rapidly extinguishing after light was turned off (fig. S2H). Finally, type 3 cells appeared to be more abundant than type 1 and 2 cells, with 30.9 cells/mm^2. Although type 1 and 2 ipRGCs' discharge rate was increased in response to the high-irradiance stimulations, their response sensitivity, latency, and duration overall were not affected by exogenous 11-*cis*-retinal (fig. S3).

To determine whether these type 3, 11-*cis*-retinal–dependent cells were specific to humans, we recorded photoresponses in retinas from adult retinal degeneration (*rd*) mice. These mice exhibit extensive degeneration of rod and cone photoreceptors; thus, light responses of 3-month-old *rd* mice are predominantly from ipRGCs (*18*). *rd* retinas produced trains of action potentials that were similar to responses recorded in human ipRGCs (fig. S4A). Upon supplementation with 11-*cis*-retinal, the number of responding cells increased (fig. S4, A and D). PCA revealed that, as in human

Fig. 3. Spectral sensitivity of human ipRGCs. (**A**) Dose–response curves of each subtype to 30-s light pulses at different irradiances and wavelengths (irr1 to irr4, from 5×10^{10} to 5×10^{14} photons/cm^2 per second; 447, 470, 505, 530, and 560 nm) (type 1, *n* = 7; type 2, *n* = 8; type 3, *n* = 79; one donor). Error bars indicate SEM. (**B**) Action spectra and (**C**) best-fitted nomograms for each subtype. EC, irradiance required to drive a 50% response; λ, wavelength.



retinas, 11-*cis*-retinal–dependent mouse ipRGCs formed a distinct cluster (fig. S4B). These cells displayed response characteristics that were different from those of other mouse ipRGCs, but similar to the type 3 cells we discovered in human retinas (fig. S4C).

To determine the spectral sensitivity of the human intrinsic light response, we measured the discharge rate as a function of wavelength (447, 470, 505, 530, and 560 nm) over ~4-log-unit irradiance range (~5×10^{11} to 5×10^{14} photons/cm^2 per second) (Fig. 3A and fig. S5). By calculating half-saturation values from the dose–response curves for each wavelength (Fig. 3A), we established the action spectra for each subtype (Fig. 3B). These data were fitted by A1 visual pigment nomograms (*19*) with peaks at 459 and 457 nm for ipRGC types 1 and 2, respectively (Fig. 3C). These responses are distinct from those of human rod, S, M, and L cone pigments (498, 420, 534, and 564 nm, respectively) but consistent with indirect measurements obtained for nocturnal melatonin suppression, pupillary reflex to light, and other melanopsin-dependent responses (table S1). Because of type 3 ipRGCs' limited range of responses, we did not fit their dose-response curves. Nevertheless, type 3 ipRGCs also appeared to be most sensitive around 470 nm (Fig. 3B).

In addition to intrinsic photosensitivity, ipRGCs also transmit rod- and cone-initiated photoresponses. To characterize rod and cone input, we compared ipRGCs' responses before and after the application of the synaptic blockers. In the absence of synaptic blockers, a large number of RGCs responded to light with conventional on- or off-type responses. These results are consistent with previous reports of light responses in the human retina, and they demonstrate that the outer, inner, and intermediate layers of the retina were functionally preserved in our preparation (*20*, *21*). After incubation with blockers and dark adaptation for 45 min, conventional RGCs stopped responding to subsequent light pulses, whereas ipRGC responses persisted (Fig. 4A). Comparison of ipRGC photoresponses before and after synaptic blockers revealed subtype-specific specialization in integrating outer retina responses with intrinsic photosensitivity (Fig. 4A and fig. S6A).

For all subtypes, extrinsic input to ipRGCs shortened response latency and lowered response thresholds. Latency was reduced by 58.4, 68.2, and 85.2% for types 1, 2, and 3, respectively, at 3.5×10^{12} photons/cm^2 per second (Fig. 4B, right, and fig. S6A); response thresholds were lowered such that ~90% of cells responded to the lowest irradiance (Fig. 4B, left). However, extrinsic input accounted for a significant portion of the sustained response and increased its sensitivity only for type 2 and 3 ipRGCs (Fig. 4C).

In summary, human ipRGCs can be separated into at least three subtypes on the basis of their responses to light, chromophore supply, and relationships with rods and cones (table S3). Different modes of chromophore regeneration within the different populations of ipRGCs may be linked to the heterogeneity in signal transduction mechanisms; in mice, a rhabdomeric cascade exists in M1 ipRGCs, a ciliary cascade operates in M4 cells, and both coexist in M2 cells (*22*).

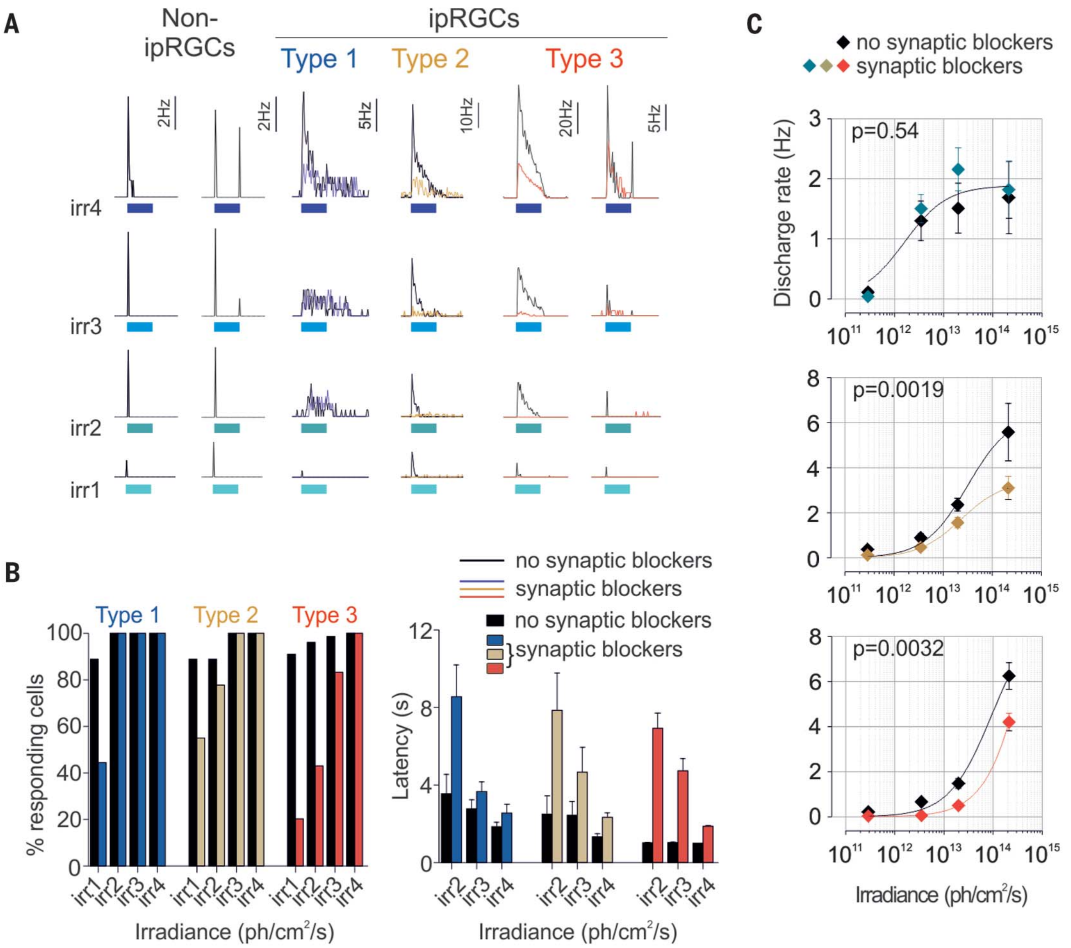


Fig. 4. Human ipRGCs integrate extrinsic signals. (**A**) Individual examples of conventional RGCs' and type 1, 2, and 3 ipRGCs' responses to increasing irradiance light pulses before (black traces) and after (color traces) application of synaptic blockers (30 s, 470 nm; irr1, irr2, irr3, and irr4: 2.9×10^{11}, 3.5×10^{12}, 2×10^{13}, and 2×10^{14} photons/cm^2 per second, respectively). Response properties before (black bars and symbols) and after (colored bars and symbols) application of synaptic blockers are shown in (B) and (C). (**B**) Threshold and latency (effect of irradiance: type 1, $P < 0.05$; type 2, $P < 0.001$; type 3, $P < 0.001$; effect of drugs: type 1, $P < 0.01$; type 2, $P < 0.01$; type 3, $P < 0.001$; two-way ANOVAs, Bonferroni post hoc tests). (**C**) Average ipRGCs' sensitivity. In both (B) and (C): type 1, $n = 9$; type 2, $n = 9$; and type 3, $n = 78$. Error bars in (C) indicate SEM.

It has long been assumed that the primary function of ipRGCs is to count photons (*14*). Our results both confirm and expand this role; rod- and cone-initiated extrinsic photosensitivity allows ipRGCs to respond to lower intensities of light and helps to sustain type 3 responses to very bright light. We also found that input from the outer retina to ipRGCs exceeds additive excitation. We observed several types of interactions—on, on or off activation, and on or off inhibition (Fig. 4A and fig. S6B)—suggesting that ipRGCs are able to mediate complex signal processing. Our study provides direct quantitative data on human ipRGC function, which will help in the development of light-based interventions for improving human health.

REFERENCES AND NOTES

1. M. T. H. Do, K.-W. Yau, *Physiol. Rev.* **90**, 1547–1581 (2010).
2. T. M. Schmidt, S.-K. Chen, S. Hattar, *Trends Neurosci.* **34**, 572–580 (2011).
3. D. C. Tu *et al.*, *Neuron* **48**, 987–999 (2005).
4. X. Zhao, B. K. Stafford, A. L. Godin, W. M. King, K. Y. Wong, *J. Physiol.* **592**, 1619–1636 (2014).
5. I. Provencio *et al.*, *J. Neurosci.* **20**, 600–605 (2000).
6. S. Nasir-Ahmad, S. C. S. Lee, P. R. Martin, U. Grünert, *J. Comp. Neurol.* **527**, 312–327 (2019).
7. H.-W. Liao *et al.*, *J. Comp. Neurol.* **524**, 2845–2872 (2016).
8. J. Hannibal, A. T. Christiansen, S. Heegaard, J. Fahrenkrug, J. F. Kiilgaard, *J. Comp. Neurol.* **525**, 1934–1961 (2017).
9. G. Esquiva, P. Lax, J. J. Pérez-Santonja, J. M. García-Fernández, N. Cuenca, *Front. Aging Neurosci.* **9**, 79 (2017).
10. K. A. Jones *et al.*, *Nat. Chem. Biol.* **9**, 630–635 (2013).
11. K. Y. Wong, *J. Neurosci.* **32**, 11478–11485 (2012).
12. K. Y. Wong, F. A. Dunn, D. M. Graham, D. M. Berson, *J. Physiol.* **582**, 279–296 (2007).
13. D. M. Berson, F. A. Dunn, M. Takao, *Science* **295**, 1070–1073 (2002).
14. D. Karnas, D. Hicks, J. Mordel, P. Pévet, H. Meissl, *PLOS ONE* **8**, e73343 (2013).
15. D. M. Dacey *et al.*, *Nature* **433**, 749–754 (2005).
16. L. E. Quattrochi *et al.*, *J. Comp. Neurol.* **527**, 297–311 (2019).
17. Y. Fu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10339–10344 (2005).
18. S. Panda *et al.*, *Science* **301**, 525–527 (2003).
19. T. D. Lamb, *Vision Res.* **35**, 3083–3091 (1995).
20. G. W. Weinstein, R. R. Hobson, F. H. Baker, *Science* **171**, 1021–1022 (1971).
21. T. W. Kraft, D. M. Schneeweis, J. L. Schnapf, *J. Physiol.* **464**, 747–765 (1993).
22. Z. Jiang, W. W. S. Yue, L. Chen, Y. Sheng, K.-W. Yau, *Cell* **175**, 652–664.e12 (2018).

ACKNOWLEDGMENTS

We thank T. Nguyen at Salk facility services for a custom-built LED array for light stimulations and D. O'Keefe for manuscript editing. **Funding:** This study was supported by NIH grants EY 016807, S10 RR027450, and NS066457 to S.P.; EY026651 and an unrestricted grant from RPB to the Ophthalmology Department at the University of Utah to F.V.; philanthropic foundation grants to A.H.; and Fondation Fyssen and Catharina Foundation fellowships to L.S.M. **Author contributions:** L.S.M., F.V., A.H., and S.P. conceived the experiments. A.H. coordinated tissue collections and harvests. L.S.M. and F.V. carried out the experiments. L.S.M. and S.P. analyzed the data. L.S.M. and S.P. wrote the manuscript. **Competing interests:** S.P. is the author of *The Circadian Code*, for which he receives author royalties. **Data and materials availability:** All data that support the findings of this study are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1251/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S3
References (*23–42*)

View/request a protocol for this paper from *Bio-protocol*.

11 August 2019; accepted 6 November 2019
10.1126/science.aaz0898

BIOCATALYSIS

Design of an in vitro biocatalytic cascade for the manufacture of islatravir

Mark A. Huffman[1]*, Anna Fryszkowska[1]*, Oscar Alvizo[2], Margie Borra-Garske[2], Kevin R. Campos[1], Keith A. Canada[1], Paul N. Devine[1], Da Duan[2], Jacob H. Forstater[1], Shane T. Grosser[1], Holst M. Halsey[1], Gregory J. Hughes[1], Junyong Jo[1], Leo A. Joyce[1]†, Joshua N. Kolev[1], Jack Liang[2], Kevin M. Maloney[1], Benjamin F. Mann[1], Nicholas M. Marshall[1]‡, Mark McLaughlin[1], Jeffrey C. Moore[1], Grant S. Murphy[1], Christopher C. Nawrat[1], Jovana Nazor[2], Scott Novick[2], Niki R. Patel[1], Agustina Rodriguez-Granillo[3]§, Sandra A. Robaire[1], Edward C. Sherer[3], Matthew D. Truppo[1]¶, Aaron M. Whittaker[1], Deeptak Verma[3], Li Xiao[3], Yingju Xu[1], Hao Yang[1]

Enzyme-catalyzed reactions have begun to transform pharmaceutical manufacturing, offering levels of selectivity and tunability that can dramatically improve chemical synthesis. Combining enzymatic reactions into multistep biocatalytic cascades brings additional benefits. Cascades avoid the waste generated by purification of intermediates. They also allow reactions to be linked together to overcome an unfavorable equilibrium or avoid the accumulation of unstable or inhibitory intermediates. We report an in vitro biocatalytic cascade synthesis of the investigational HIV treatment islatravir. Five enzymes were engineered through directed evolution to act on non-natural substrates. These were combined with four auxiliary enzymes to construct islatravir from simple building blocks in a three-step biocatalytic cascade. The overall synthesis requires fewer than half the number of steps of the previously reported routes.

The structural complexity of new drug molecules and the growing desire to develop green and efficient synthetic processes demand innovation and excellence in organic chemistry (*1*). Enzyme catalysis incorporated into pharmaceutical manufacturing represents one such innovation, providing benefits that include unparalleled selectivity, increased atom economy, and improved safety (*2*, *3*). But a truly transformative potential lies in combining two or more enzymatic steps into biocatalytic cascade sequences (*4*, *5*). Biocatalytic cascades save resources by avoiding isolation of intermediates. They also allow thermodynamically unfavorable steps to be coupled to more favorable reactions and can avoid enzyme inhibition by consuming inhibitory intermediates as they are formed. Cascades are enabled by the exceptional chemoselectivity of enzymes and their compatibility with a common set of mild aqueous reaction conditions. Designing chemical syntheses around the use of multienzyme cascades could revolutionize the manufacture of drugs. Putting this vision into practice requires a capacity for rapid identification and engineering of multiple enzymes to act on unnatural substrates at industrially relevant concentrations. Ongoing advances in directed evolution (*6*, *7*), combined with an increasing abundance of genomic information, have now brought this goal within reach, enabling us to construct a nine-enzyme cascade to manufacture the investigational HIV treatment islatravir.

[1]Process Research and Development, Merck & Co., Inc., Rahway, NJ 07065, USA. [2]Codexis, Inc., 200 Penobscot Drive, Redwood City, CA 94063, USA. [3]Computational and Structural Chemistry, Discovery Chemistry, Merck & Co., Inc., Kenilworth, NJ 07033, USA.
*Corresponding author. Email: mark_huffman@merck.com (M.A.H.); anna.fryszkowska@merck.com (A.F.) †Present address: Arrowhead Pharmaceuticals, 502 South Rosa Road, Madison, WI 53719, USA. ‡Present address: Invenra, 505 South Rosa Road, Madison, WI 53719, USA. §Present address: Schrodinger, 222 Third St, Cambridge, MA 02142, USA. ¶Present address: Janssen Research & Development, Spring House, PA 19477, USA.

The nucleoside analog islatravir (MK-8591, EFdA, **1**) is an HIV reverse transcriptase translocation inhibitor with a previously undeveloped mechanism of action (*8*, *9*). Its extraordinary potency and long duration of action promise utility in reduced-frequency dosing regimens for HIV treatment and preexposure prophylaxis (*10*). Several synthetic routes to islatravir have been published, each requiring between 12 and 18 steps (*11*–*15*). Inefficiencies in the previous syntheses arose from the need for multiple protecting-group manipulations as well as the difficulty of controlling anomeric stereochemistry in 2′-deoxyribonucleosides. Enzymatic reactions often eliminate the need for protecting groups while at the same time providing precise control over stereoselectivity. We therefore envisaged that biocatalysis might effectively address the synthetic challenges of islatravir.

The bacterial nucleoside salvage pathway (Fig. 1A) provides an attractive biocatalytic retrosynthetic scheme for deoxyribonucleosides (*16*, *17*). The biological sequence degrades purine 2′-deoxyribonucleosides using three enzymes (*18*). First, purine nucleoside phosphorylase (PNP) displaces the nucleobase with phosphate to give deoxyribose 1-phosphate. Phosphopentomutase (PPM) then transfers the phosphate to the 5 position. The resulting sugar 5-phosphate becomes a substrate for deoxyribose 5-phosphate aldolase (DERA), which performs a retro-aldol cleavage into glyceraldehyde 3-phosphate and acetaldehyde. This sequence can rapidly assemble nucleosides from simple starting materials when run in reverse (*17*). Using the salvage pathway to synthesize islatravir requires all three enzymes to act on non-natural substrates bearing a fully substituted carbon at C-4 of the 2-deoxyribose ring. The full sequence of enzymes has not been used to both construct a non-natural sugar and attach a nucleobase. Nonetheless, we were encouraged by previous applications of PNP and PPM to produce non-natural nucleosides (*17*, *19*) as well as the ability of DERA to make non-natural sugars (*20*).

The reversibility of the reactions in the salvage pathway allowed us to rapidly explore the sequence in the retrosynthetic direction (*19*), starting from islatravir. We began by evaluating a panel of purine phosphorylases for their ability to cleave islatravir into the nucleobase (**2**) and sugar 1-phosphate (**3**). The ethynyl substituent was accepted by several homologs, with the native *Escherichia coli* PNP showing the highest activity. Application of the *E. coli* enzyme at high loading generated the otherwise inaccessible 1-phosphate intermediate (**3**) to use in testing phosphopentomutases. The native *E. coli* PPM displayed the highest activity for converting this unnatural substrate to the thermodynamically favored 5-phosphate (**4**). With compelling proof of concept for the two desired reactions in place, we then focused on improving the activity of both enzymes by means of directed evolution.

We used a homology model based on the *Bacillus cereus* PPM crystal structure to generate a library of variants with single mutations near the active site, which were screened for improved activity in the reverse reaction. In the best performing variant, the manganese-binding D172 was replaced by a noncoordinating alanine (fig. S1D). Recombination of beneficial mutations from the first round—including a change in the catalytic, phosphorous-transferring residue T97S—resulted in a quintuple mutant that exhibited ≥70-fold improvement over the wild-type enzyme (Table 1 and figs. S22 to S24). (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, T97S indicates that threonine at position 97 was replaced by serine.)

The improved PPM variant enabled us to evolve PNP for improved activity. We screened a library of variants with single mutations near the active site for formation of islatravir in

the tandem forward reaction from sugar 5-phosphate (**4**), with the 1-phosphate (**3**) being generated in situ by the newly evolved PPM. A single amino acid change, M64A, near the alkyne-bearing C-4 of the docked substrate (fig. S1E) provided a PNP variant with about 350-fold improved activity (Table 1 and figs. S26 and S27). With the two active enzymes, we focused our attention on evaluating the practicality of the glycosylation cascade.

Because the PPM-catalyzed equilibrium favors the starting 5-phosphate, the phosphate transfer and glycosylation reactions must be performed simultaneously. Despite the improved activity of both enzymes, the tandem reaction plateaued at <50% yield. A reverse reaction starting from islatravir reached 40% conversion, demonstrating that the reaction was equilibrium limited. In addition, the inorganic phosphate byproduct of the glycosylation reaction is known to inhibit PPM (*21*). Removing the inorganic phosphate as it is formed provided an effective solution to the equilibrium and inhibition problems (*22*). Addition of sucrose phosphorylase (SP) and sucrose to the reaction converts free phosphate to glucose 1-phosphate and shifts the entire equilibrium forward. The resulting three-enzyme cascade runs to full conversion at substrate concentrations as high as 200 mM (scheme S4B).

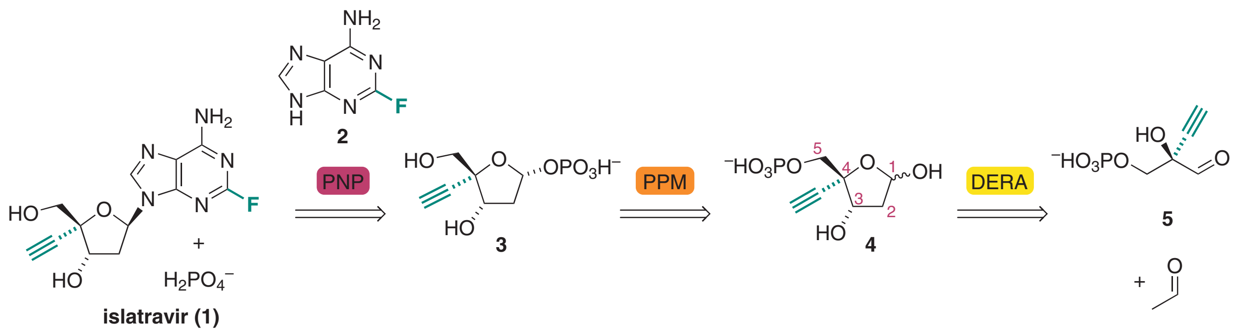
We continued exploring the retrosynthetic degradation pathway by screening DERA enzymes in the retro-aldol reaction. Several DERA homologs displayed high activity in the retro-aldol of the alkynylated sugar. After further evaluation in the forward aldol reaction, we selected the DERA from *Shewanella halifaxensis* for its high activity, complete stereoselectivity in the formation of the new C–C bond (> 99% de), and kinetic selectivity favoring reaction with the (*R*)-enantiomer of the aldehyde. The (*S*)-aldehyde was very slowly converted to the (*3S,4S*) diastereomer. This wild-type DERA was also active in reaction with the nonphosphorylated aldehyde (**7**). Despite extensive research on applications of DERA enzymes, we did not find any reports of reactions of aldehydes with a fully substituted α-carbon.

The limitation of the *S. halifaxensis* DERA was that it did not tolerate the high concentrations of acetaldehyde required for a practical synthesis. Two rounds of directed evolution addressed this constraint, yielding an improved variant that retained high activity at an acetaldehyde concentration >400 mM (Table 1). The engineered enzyme contained several new mutations, including C47V (fig. S1C). This cysteine residue is found near the active site in other DERA homologs and acts as a regulator of the enzyme activity, binding the product of acetaldehyde self-aldol condensation (*23*).

With the repurposed nucleoside salvage enzymes in hand, we focused on an efficient synthesis of 2-ethynylglyceraldehyde 3-phosphate (**5**). Known biochemical pathways to its natural analog (*R*)-glyceraldehyde phosphate do not allow for facile incorporation of a C-2-substitutent. Retrosynthetic analysis suggested that **5** could be accessed from the simple achiral building block 2-ethynylglycerol (**6**) through oxidation and phosphorylation reactions (Fig. 1B). In theory, these transformations could occur in either order. We explored enzymes for all four of these reactions and did not find any with activity for oxidation of **8**. We therefore pursued the path of oxidation followed by phosphorylation.

Discovery of an enzyme with phosphorylation activity toward **7** required extensive screening of a broad spectrum of kinases that naturally act on sugars and primary alcohols. Ultimately, we identified a pantothenate kinase (PanK) from *E. coli* with very low-level activity toward the (*R*)-enantiomer of aldehyde **7**. Applying directed evolution, we were able to rapidly increase the activity by introducing two mutations (L277I and I281M) in the helix of the pantothenate binding site (fig. S1B). Further engineering improved the enzyme's activity and stability. The evolved variant displayed greater than 100-fold higher activity (fig. S19) and pro-(*R*) kinetic selectivity ($E = 8$) (scheme S19), achieving full conversion at 200 mM concentration (Table 1). Practical use of the kinase in vitro required regeneration of its cofactor adenosine triphosphate. We chose

A **Bacterial nucleoside salvage pathway applied to islatravir (1).** *Unnatural structural elements are highlighted in green.*



B **Biocatalytic retrosynthetic analysis of aldehyde intermediate (5).**

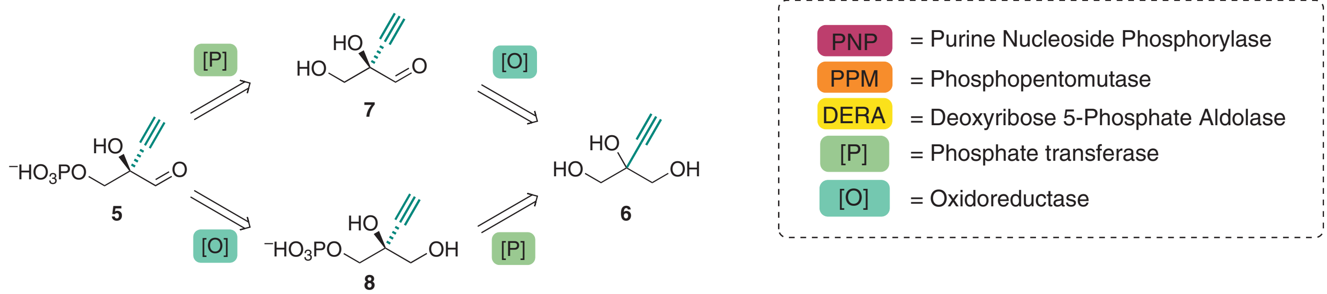


Fig. 1. Biocatalytic retrosynthetic planning. (**A**) Purine nucleoside degradation pathway applied retrosynthetically to islatravir (**1**). (**B**) Retrosynthesis of the glyceraldehyde phosphate analog leading to a simple achiral building block.

acetyl phosphate as an economical source of activated phosphate paired with a thermostable acetate kinase (AcK) from *Thermotoga maritima* (Fig. 2).

The final piece to complete the biocatalytic pathway was the desymmetrizing oxidation of 2-ethynylglycerol (**6**). Among a broad range of oxidoreductases, we identified evolved variants of galactose oxidase (GOase M1 and F2) that chemoselectively formed the monoaldehyde **7** with limited overoxidation (Table 1). These copper-dependent enzymes were previously engineered for improved expression in *E. coli* (M1-variant) (*24*) and broadened substrate scope (F2-variant) (*25*). Unfortunately, both M1 and F2 GOase variants favored formation of the undesired (*S*)-enantiomer with 8:92 and 40:60 *R*:*S* ratios, respectively. The (*S*)-selectivity is consistent with the stereochemistry of the natural substrate D-galactose. Nevertheless, the relaxed stereoselectivity of the F2-variant suggested that its enantiopreference could be reversed, and the GOase F2 became the starting point for evolution.

Synthetic applications of GOases require two additional redox enzymes: a catalase to disproportionate the hydrogen peroxide byproduct and a peroxidase to maintain the correct oxidation state of copper (*26*). We chose commercially available bovine catalase and horseradish peroxidase and used these enzymes during evolution of the GOase. Over 12 rounds of oxidase engineering, we targeted an improvement in activity, a reduction in product inhibition, and a reversal of the innate enantiopreference. The engineered oxidase displayed 11-fold improved activity and a reversed 90:10 *R*:*S* selectivity (Table 1). The change in enantioselectivity resulted mainly from two mutations in the active site: W290Y and F464L (fig. S1A). Multiple additional mutations throughout the enzyme improved protein expression, stability, and activity and alleviated product inhibition. With higher activity came an increase in overoxidation of the product. The overoxidation led to an increase in the enantiopurity of **7** throughout the course of the reaction (up to 99% enantiomeric excess), albeit at the cost of yield (fig. S10).

At this point, we had evolved enzymes for each step of a fully biocatalytic sequence from ethynyl glycerol (**6**) to islatravir. We then turned to strategic considerations around how best to integrate these individual reactions into the overall in vitro synthesis. As discussed above, the three reversible reactions catalyzed by PPM, PNP, and SP must take place concurrently to achieve a favorable equilibrium. Extending the simultaneous cascade to include the reversible aldol addition could provide additional synergy. Compatible pH and temperature ranges would allow all four enzymes to function in the same solution. Sucrose phosphorolysis as a thermodynamic driving force could pull forward all the equilibria, including the aldol reaction. These factors enabled us to lower the excess of acetaldehyde to levels tolerated by all the enzymes. In this way, we could operate a simultaneous four-enzyme cascade in which the nucleoside degradation pathway runs in reverse, driven to high conversion by phosphate removal (Fig. 2). Islatravir crystallized from the reaction mixture and could be isolated directly through filtration in greater than 95% purity and 76% yield from **5**.

The oxidation and phosphorylation reactions are essentially irreversible, so directly coupling them with the downstream cascade provides no thermodynamic benefit. We considered a tandem GOase-kinase combined reaction, so that in situ product consumption would minimize the inhibition suffered by the oxidase. Unfortunately, the pantothenate kinase lacked the required chemoselectivity, rapidly phosphorylating triol **6**.

Isolating the intermediates **5** and **7** proved to be challenging because of their high solubility in water. Therefore, we focused on developing a process in which a single aqueous solution was carried through the entire sequence. This approach necessitated evaluating the impact of each component on the remaining steps. The use of acetyl phosphate as the phosphorous donor requires neutralization of the liberated acetic acid, generating a solution with elevated ionic strength. The high salt content inhibits downstream enzymes—in particular, the hexameric PNP (figs. S15 and S16). To enable these reactions to proceed effectively, we carried out further PNP evolution under high-salt conditions and diluted the kinase product solution to reduce its ionic strength before the next step.

With a total of nine enzymes used in the synthesis and no intermediate isolations, management of cumulative protein content was critical. The final filtration of crystallized islatravir became difficult as the amount of protein in the mixture increased. The homogeneous nature of the oxidation and phosphorylation reactions opened the possibility of immobilizing their enzymes to allow easy removal after these steps. For this purpose, we applied an affinity immobilization technique that relies on the capture of polyhistidine-tagged proteins on a metal-containing solid support (*27*). The recombinant oxidase and kinase enzymes were immobilized in this way, which had the added benefit of eliminating nontagged proteins during immobilization and minimizing

Table 1. Properties and performance of evolved enzymes used in the biocatalytic pathway.

					Starting enzyme		Evolved variant	
Enzyme	Source organism	Evolution focus	Rounds of evolution	Global amino acids changed (no.)	Conversion (selectivity)	Loading (%)*	Conversion (selectivity)	Loading (%)*
Oxidase (GOase)	*Fusarium graminearum*	Stereoselectivity Activity	12	34	Variant M1:33%†‡ (8:92 *R*:*S*) Variant F2: 8%†‡ (40:60 *R*:*S*)	100	80%†§ (90:10 *R*:*S*)	20
PanK	*E. coli*	Activity	3	10	<1% (5:1 *R*:*S*)†¶	10	>95% (10:1 *R*:*S*)†#	10
DERA	*S. halifaxensis*	Acetaldehyde tolerance	2	11	97%**[>98:1:1 (3*S*4*R*):(3*R*4*R*):(3*S*4*S*)]	5	97%**[>98:1:1 (3*S*4*R*):(3*R*4*R*):(3*S*4*S*)]	0.2
PPM	*E. coli*	Activity	2	5	0.5%††	0.5	34%††	0.5
PNP	*E. coli*	Activity	4	7	0.18%‡‡ (>99.5:0.5 *dr*)	0.125	62%‡‡ (>99.5:0.5 *dr*)	0.125

*Enzyme loading refers to the mass of lyophilized clarified cell lysate relative to the mass of the reaction substrate. Results may reflect improvements in enzyme expression as well as activity. †Reaction with nonimmobilized enzymes. ‡GOase-M1-Strep and F2-Strep: 172 mM **6**, pH 7.5, 0.2 mM CuSO4, 25°C, 4 hours. §GOAse-13BB-His: 258 mM **6**, pH 7.5, 0.2 mM CuSO4, 25°C, 4 hours. ¶44 mM **7**, pH 7.5, 20°C, 18 hours. #235 mM **7**, pH 6.4, 20°C, 18 hours. **142 mM **5**, 420 mM acetaldehyde, pH 7.2, 30°C, 24 hours. ††15 mM **4**, 5 mM $MnCl_2$, pH 7.5, 40°C, 18 hours. ‡‡13 mM **1b**, pH 7.5, 40°C, 16 hours.

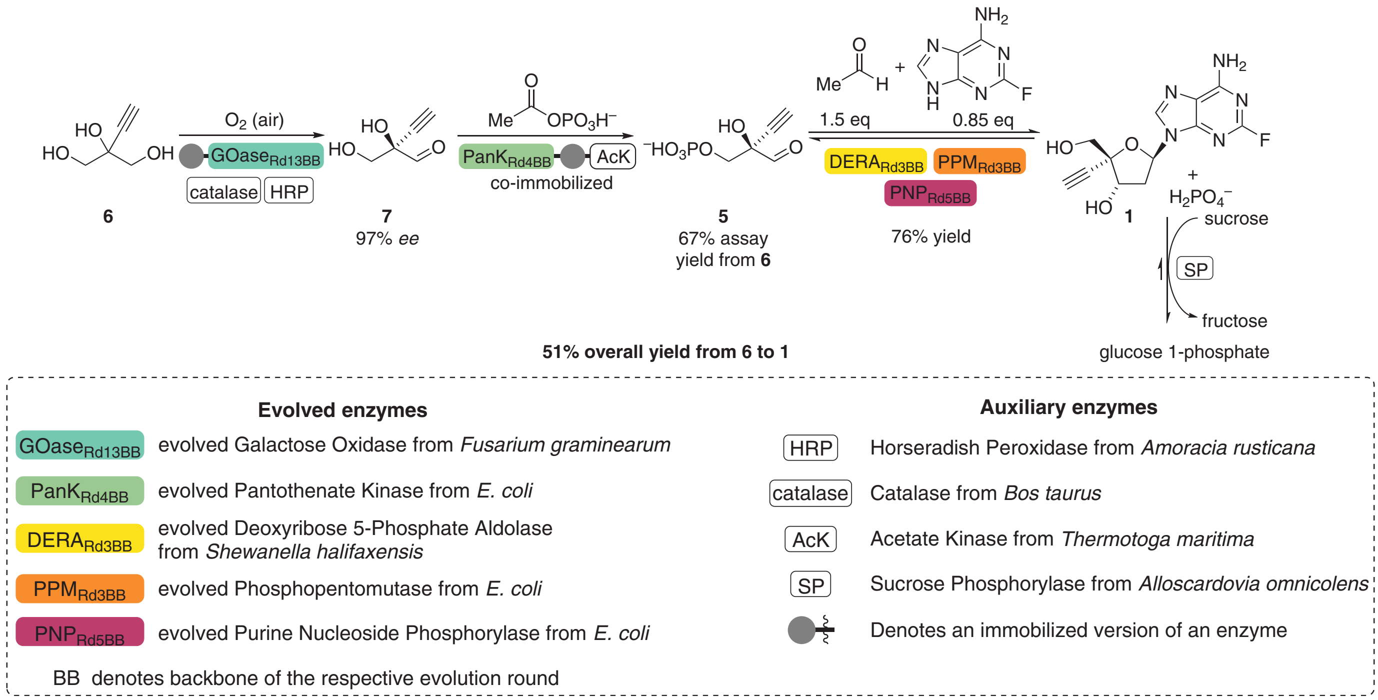


Fig. 2. Fully assembled biocatalytic pathway. Evolved enzymes are in colored boxes, and wild-type auxiliary enzymes are in white boxes.

undesired side reactions catalyzed by host-cell enzymes. Immobilization was not pursued for the aldol and glycosylation enzymes because the product crystallizes during this step.

The biocatalytic synthesis produces only a single stereoisomer of islatravir in a self-correcting manner. Oxidative desymmetrization establishes the fully substituted carbon center with 90:10 selectivity, and the ratio is increased by further oxidation of the minor enantiomer. Every subsequent step provides an opportunity to amplify the stereochemical purity. The kinase and aldolase enzymes react with kinetic selectivity toward the (*R*)-aldehydes **7** and **5**, respectively, allowing upgrade of the enantiomer ratio. The aldolase creates the second stereogenic center with great precision. Last, only the (3*S*,4*R*)-diastereomer of the sugar phosphate **4** can react further, and the glycosylation sets the anomeric center with perfect selectivity.

The full in vitro biocatalytic cascade uses five engineered enzymes and four auxiliary enzymes to stereoselectively assemble islatravir from simple achiral building blocks in 51% overall yield. The atom economy far exceeds that of previous syntheses of this target, and the number of steps is less than half. The entire sequence takes place under mild conditions in a single aqueous solution without the isolation of intermediates. This extraordinary efficiency was made possible by the ability to identify and engineer enzymes that can build complex structures with excellent stereo- and chemoselectivity and without the need for protecting groups. We envision a growing adoption of cascade biocatalysis as a strategy for the sustainable synthesis of complex non-natural molecules such as pharmaceuticals. The application of enzyme cascades to a diverse range of molecular structures will rely on further advances in the pace of protein engineering (*28*) and the continuing discovery of new enzymatic transformations (*7*, *29*).

REFERENCES AND NOTES

1. K. R. Campos *et al.*, *Science* **363**, eaat0805 (2019).
2. R. A. Sheldon, J. M. Woodley, *Chem. Rev.* **118**, 801–838 (2018).
3. N. J. Turner, L. Humphreys, *Biocatalysis in Organic Synthesis* (The Royal Society of Chemistry, 2018).
4. J. H. Schrittwieser, S. Velikogne, M. Hall, W. Kroutil, *Chem. Rev.* **118**, 270–348 (2018).
5. S. P. France, L. J. Hepworth, N. J. Turner, S. L. Flitsch, *ACS Catal.* **7**, 710–724 (2017).
6. F. H. Arnold, *Q. Rev. Biophys.* **48**, 404–410 (2015).
7. F. H. Arnold, *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).
8. H. Ohrui *et al.*, *Nucleosides Nucleotides Nucleic Acids* **26**, 1543–1546 (2007).
9. E. Michailidis *et al.*, *J. Biol. Chem.* **289**, 24533–24548 (2014).
10. S. E. Barrett *et al.*, *Antimicrob. Agents Chemother.* **62**, e01058-18 (2018).
11. M. McLaughlin *et al.*, *Org. Lett.* **19**, 926–929 (2017).
12. M. Kageyama, T. Nagasawa, M. Yoshida, H. Ohrui, S. Kuwahara, *Org. Lett.* **13**, 5264–5266 (2011).
13. K. Fukuyama, H. Ohrui, S. Kuwahara, *Org. Lett.* **17**, 828–831 (2015).
14. M. Kageyama *et al.*, *Biosci. Biotechnol. Biochem.* **76**, 1219–1225 (2012).
15. S. Kohgo, H. Ohrui, M. Matsuoka, H. Mitsuya, 4′-C-substituted-2-haloadenosine derivative, U.S. patent no. US7625877 B2 (2009).
16. N. J. Turner, E. O'Reilly, *Nat. Chem. Biol.* **9**, 285–288 (2013).
17. I. A. Mikhailopulo, A. I. Miroshnikov, *Acta Naturae* **2**, 36–59 (2010).
18. M. G. Tozzi, M. Camici, L. Mascia, F. Sgarrella, P. L. Ipata, *FEBS J.* **273**, 1089–1101 (2006).
19. W. R. Birmingham *et al.*, *Nat. Chem. Biol.* **10**, 392–399 (2014).
20. C. L. Windle, M. Müller, A. Nelson, A. Berry, *Curr. Opin. Chem. Biol.* **19**, 25–33 (2014).
21. K. Hammer-Jespersen, A. Munch-Petersen, *Eur. J. Biochem.* **17**, 397–407 (1970).
22. W. Tischer, H.-G. Ihlenfeldt, O. Barzu, H. Sakamoto, E. Pistotnik, P. Marlière, S. Pochet, Enzymatic synthesis of deoxyribonucleosides, U.S. patent no. US07229797 B1 (2007).
23. M. Dick *et al.*, *Chem. Sci.* **7**, 4492–4502 (2016).
24. L. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara, F. H. Arnold, *Protein Eng. Des. Sel.* **14**, 699–704 (2001).
25. J. B. Rannes *et al.*, *J. Am. Chem. Soc.* **133**, 8436–8439 (2011).
26. A. Toftgaard Pedersen *et al.*, *Org. Process Res. Dev.* **19**, 1580–1589 (2015).
27. M. P. Thompson *et al.*, *Tetrahedron* **75**, 327–334 (2019).
28. M. D. Truppo, *ACS Med. Chem. Lett.* **8**, 476–480 (2017).
29. P. N. Devine *et al.*, *Nat. Rev. Chem.* **2**, 409–421 (2018).

ACKNOWLEDGMENTS

We acknowledge the help and support of the following people. Merck team: J. McIntosh for providing wild-type acid phosphatase and acetate kinase constructs; R. Cohen, X. Wang, M. Reibarkh, and P. Dormer for nuclear magnetic resonance analysis support; R. Patel, C. Mastykarz, W. Pan, J. Gouker, I. Farasat, J. Russell, and L. Do for supporting protein engineering and directed evolution workflows; A. Kassim, T. Andreani, and R. Matthew for help with the preparation of synthetic intermediates; and E. Margelefsky, K. Mattern, M. Miller, and H. Rose for reaction optimization support in scale-up experiments. Codexis team: A. Ortega, A. Sowell-Kantz, H. Maniar, J. Slaton, C. Micklitsch, L. Miller, and M. Krawczyk for library screening support; V. Mitchell, C. Selim, and A. Petkova for library construction support; N. Subramanian and N. Dellas for library design support; J. Riggins for analytical development support; J. Vroom and S. Sivaramakrishnan for enzyme characterization support; and D. Entwistle for chemistry support. **Funding:** This work was funded by Merck & Co., Inc. **Author contributions:** M.A.H. and A.F. carried out and supervised chemistry development and prepared the manuscript. K.R.C., P.N.D., K.M.M, and M.D.T. supervised chemistry development. O.A. and K.A.C. supervised enzyme evolution. J.H.F., S.T.G., G.J.H., J.N.K., J.C.M., M.M, C.C.N., N.R.P., S.A.R., A.M.W., Y.X., and H.Y. carried out chemistry development. M.B.-G., D.D., J.L., N.M.M., G.S.M., J.N., and S.N. carried out enzyme evolution. H.M.H., J.J., L.A.J., and B.F.M. developed analytical methods to allow chemistry development

and enzyme evolution. A.R.-G, E.C.S., D.V., and L.X. carried out computational work to support enzyme discovery and evolution. **Competing interests:** Merck & Co., Inc. has filed a patent application covering the process: U.S. patent application no. PCT/US2019/040316. Codexis, Inc. has filed patent applications covering the evolved enzymes: U.S. patent application nos. PCT/US2019/040353, PCT/US2019/040359, PCT/US2019/040369, PCT/US2019/040376, and PCT/US2019/040379. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1255/suppl/DC1
Materials and Methods
Figs. S1 to S28
Tables S1 to S15
Schemes S1 to S81
DNA and amino acid sequences of the enzymes
NMR Spectra
HR-MS Spectra
References (*30*–*50*)

12 August 2019; accepted 24 October 2019
10.1126/science.aay8484

STRUCTURAL BIOLOGY

Structures of the AMPA receptor in complex with its auxiliary subunit cornichon

Terunaga Nakagawa

In the brain, AMPA-type glutamate receptors (AMPARs) form complexes with their auxiliary subunits and mediate the majority of fast excitatory neurotransmission. Signals transduced by these complexes are critical for synaptic plasticity, learning, and memory. The two major categories of AMPAR auxiliary subunits are transmembrane AMPAR regulatory proteins (TARPs) and cornichon homologs (CNIHs); these subunits share little homology and play distinct roles in controlling ion channel gating and trafficking of AMPAR. Here, I report high-resolution cryo–electron microscopy structures of AMPAR in complex with CNIH3. Contrary to its predicted membrane topology, CNIH3 lacks an extracellular domain and instead contains four membrane-spanning helices. The protein-protein interaction interface that dictates channel modulation and the lipids surrounding the complex are revealed. These structures provide insights into the molecular mechanism for ion channel modulation and assembly of AMPAR/CNIH3 complexes.

AMPA-type ionotropic glutamate receptors (AMPARs), ligand-gated ion channels activated by the neurotransmitter glutamate, mediate the majority of fast excitatory synaptic transmission in the brain (*1*). They regulate synaptic plasticity, which in turn influences circuit activity, cognition, learning, and behavior. Their dysfunction is associated with a variety of neurological and psychiatric disorders (*2*), including major depressive disorder, Alzheimer's disease, Rasmussen's encephalitis, limbic encephalitis, seizures, cognitive dysfunction, and autism.

In mammals, the pore-forming subunits of AMPARs are called GluA1 to -4. The extracellular domains of each subunit consist of an N-terminal domain (NTD) and a ligand-binding domain (LBD) (Fig. 1A). In the resting state, NTDs are stacked on top of the LBD and extend away from the membrane (*3*). Glutamate binds to the LBD and induces a conformational change in the transmembrane domain (TMD) that in turn triggers gating (*4*). The TMD forms the ion channel pore made of three membrane-spanning segments (M1, M3, and M4) and a re-entrant element (M2). A short cytoplasmic domain (CTD) extends into the cytoplasm. AMPARs function as tetramers and adopt a dimer-of-dimers architecture resembling a Y shape, in which the extracellular NTD and LBD each form dimers exhibiting an overall twofold symmetry, whereas the membrane-embedded portion adopts a near-fourfold symmetry (*5*) (see organization of the GluA2 tetramer in Fig. 1C). In addition, certain heterotetrameric AMPARs adopt a compact global conformation that differs from the canonical Y shape (*6*).

Department of Molecular Physiology and Biophysics, Center for Structural Biology, and Vanderbilt Brain Institute, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.
Email: terunaga.nakagawa@vanderbilt.edu

AMPARs form complexes with various structurally unrelated auxiliary subunits, which are membrane proteins that regulate AMPAR trafficking or gating (and in some cases both) (*7*). The two major classes of AMPAR auxiliary subunits belong to the claudin homolog and cornichon homolog (CNIH), which share little homology (*8*). Transmembrane AMPAR regulatory proteins (TARPs) are members of the claudin homolog family and are the most extensively studied (*7–9*). Among the cornichon family, CNIH2 and -3 (CNIH2/3) are AMPAR auxiliary subunits (*10*). Unlike the TARPs, CNIH2/3 function at the endoplasmic reticulum where, in mammals, they may control assembly of heteromeric AMPARs (*11*, *12*). CNIH2/3 remain associated with the synaptic AMPAR complex and, in many cases, coassemble with TARPs (*11*, *13*, *14*). Both TARPs and CNIH2/3 slow ion channel desensitization to varying degrees (*7*). The role of auxiliary subunits in tuning ion channel gating kinetics is predicted to have a substantial impact on circuit dynamics (*7*). Knowledge of the modulation mechanisms of AMPAR gating and trafficking used by various auxiliary subunits could guide rational design of new therapeutic compounds. Currently, our structural knowledge of AMPAR auxiliary subunit complexes has been limited to those that contain either TARPs or GSG1L, which are both claudin homologs (*15*, *16*).

I investigated cryo–electron microscopy (cryo-EM) structures of complexes composed of GluA2 and CNIH3 (hereafter referred to as A2-C3) bound to the antagonist ZK200775 (280 μM). Detailed methods on sample preparation and data collection are given in the supplementary materials. In the first step of the analyses, I obtained full-length structures at an overall resolution of 4.4 Å (figs. S3 and S4). The low flexibility between NTD and LBD interfered with further high-resolution analyses, and therefore in the second step the NTD layer and the rest of the complex (hereafter referred to as LBD-TMD-C3) were analyzed separately by focused classification and refinement (supplementary text and figs. S3 to S10). High-resolution maps, whose overall resolutions ranged from 3.0 to 3.5 Å, and their molecular models (for statistics used, see table S1) were placed into the 4.4-Å-resolution full-length map to reconstruct the complete structure (Fig. 1).

I observed two global conformations. The pseudosymmetric conformation (PS) resembles the canonical Y shape (Fig. 1C), whereas the asymmetric conformation (AS) exhibits a tilted NTD layer relative to the LBD layer and makes interlayer contact at one corner (Fig. 1, B to D). Similar numbers of particles (AS, 218,413 particles; PS, 190,470 particles) contributed to each conformation. The maps of PS and AS revealed four extra densities in the micelle attached to the TMD of GluA2; these detected densities are CNIH3, bound at a GluA2:CNIH3 stoichiometry of 4:4 (Fig. 1, B to D).

The architectures of NTD tetramers were virtually identical between AS and PS [root mean square deviation (RMSD) of Cα, 0.374 Å] and similar to previous structures (fig. S11). Glycosylation at N241, which was eliminated by mutation in previous studies (*5*, *17*), points toward LBD, potentially biasing the NTD layer tilted in AS and preventing it from descending vertically (Fig. 1, B, C, F, and G). In AS, K188 is close to I459 and Y469 at the interdomain contact between the NTD and LBD (Fig. 1, B and J). The NTDs might alter dynamics of LBDs through direct interaction, which could potentially result in allosteric gating modulation, such as that seen in *N*-methyl-D-aspartate receptors (*18*). The LBDs are nearly identical between AS and PS (RMSD of Cα, 0.542 Å) and also similar to those of GluA2/stargazin and GluA2/GSG1L in the closed state (*15*, *19*) (fig. S12). Consistently, architectures below the NTD layer were virtually identical between AS and PS (RMSD of Cα, 0.526 Å) (fig. S7E). Switching between AS and PS must be a probabilistic process that requires the native NTD-LBD linker. The pore is closed, and the lower part of the channel is more compacted than the GluA2 tetramer with no auxiliary subunit (PDB: 3KG2). This compaction geometrically occludes the M2, which was unresolved (fig. S13) (see supplementary text).

An atomic model of CNIH3 was built de novo and produced the first molecular view of a member of the cornichon family. Previous studies had relied on computational models that predict CNIHs having their N termini in the cytoplasm and spanning the membrane only three times (*8*, *10*, *20–22*). Our data, however, redefine the topology of CNIH3 and reveal a geometry resembling four transmembrane segments with extracellular N and

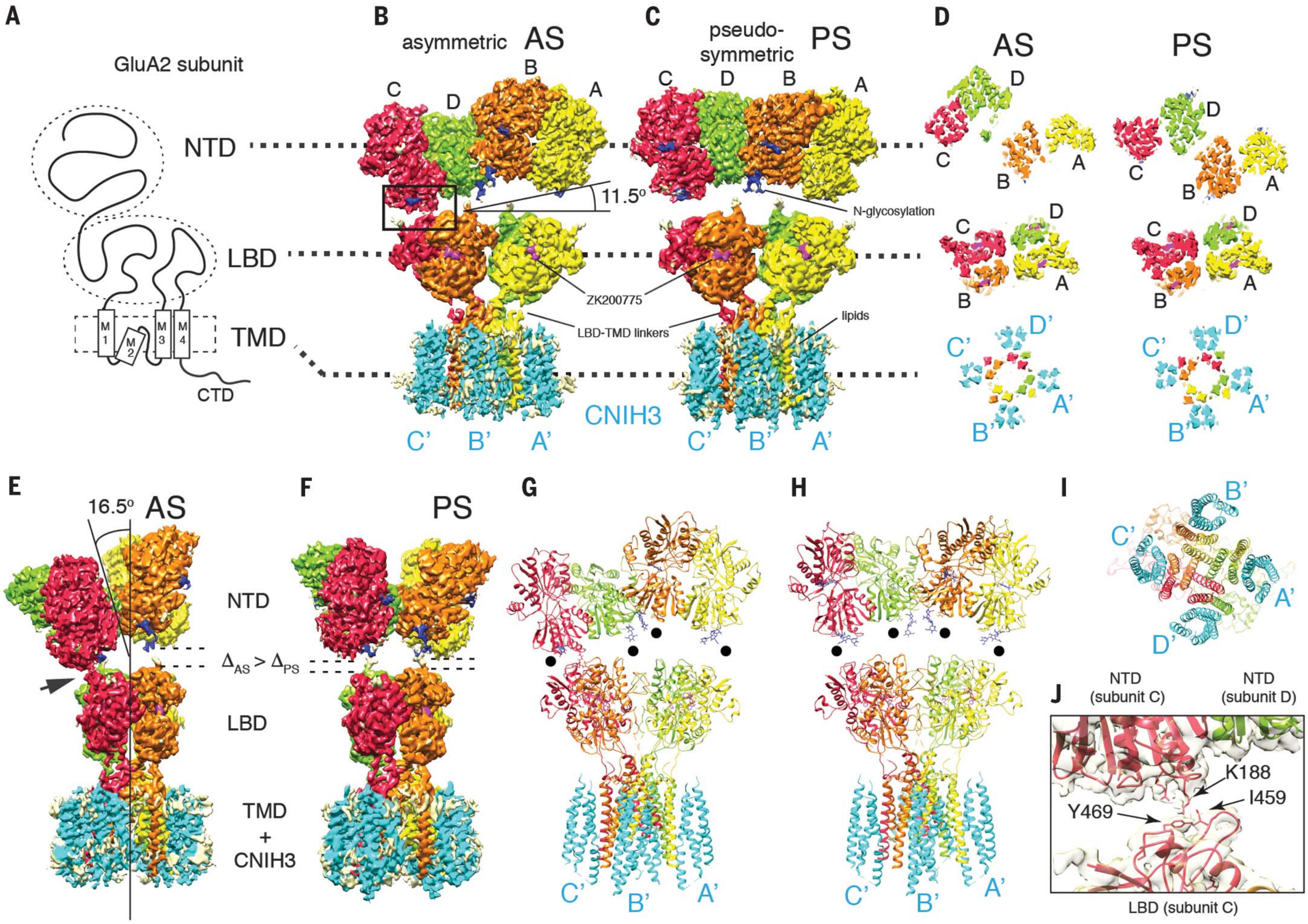


Fig. 1. Cryo-EM structures of the complex formed of GluA2 and CNIH3 (A2-C3). (**A**) Domain organization of a GluA2 subunit. (**B** and **C**) Density map of A2-C3 in AS and PS. From this view, the NTD layer is tilted at 11.5° in AS. No symmetry was imposed in solving AS, whereas C2 was imposed for PS. Visualizing thresholds: NTD(AS) at 7.07σ, LBD-TMD-C3(AS) I at 7.56σ, NTD(PS) at 7.32σ, LBD-TMD-C3(PS) at 6.80σ. Overall resolutions: NTD(AS), 3.1 Å; LBD-TMD-C3(AS) I, 3.5 Å; NTD(PS), 3.1 Å; LBD-TMD-C3, 3.2 Å. See table S1 and figs. S3 to S8 for detailed descriptions of each of the maps. (**D**) Cross sections of each domain (indicated by dashed lines), viewed from the top (the NTD side). The subunits of tetrameric GluA2 are referred to as A (yellow), B (orange), C (red), and D (green). The cyan densities are CNIH3, named A′ to D′ based on location, following the style used for TARPs (*16*, *25*). (**E** and **F**) Side views of maps shown in (B) and (C). The tilt angle of the NTD layer (16.5°), the NTD-LBD contact (arrow), and gaps between NTD and LBD (Δ_{AS}, Δ_{PS}) are indicated. (**G** to **I**) Molecular models of A2-C3 in AS and PS, shown as ribbon diagrams. Models were built from maps NTD(AS), LBD-TMD-C3(AS)II, NTD(PS), and LBD-TMD-C3(PS) (table S1). Black dots indicate glycosylation at N241. The bottom view is shown in (I). (**J**) Zoomed-in view of the NTD-LBD contact in the C subunit, indicated as a rectangle in (B). Model and map [NTD(AS) at 7.07σ] are superimposed.

C termini (Fig. 2, A and B). CNIH3 lacks a canonical signal peptide, but the N terminus, which remains uncleaved and buried in the membrane, appears to substitute for it. The first 12 amino acids are almost identical in mammals, flies, and worms, but not in plants or yeast (Fig. 2C). I refer to this fragment as uncleavable membrane inserting peptide (UMIP). UMIP of CNIH2 is intolerant to missense mutations in humans (*23*) (fig. S14).

The TM1 helix begins within UMIP, and the end of the helix penetrates into the cytoplasm. After a short unresolved loop, the TM2 starts as a cytoplasmic helix, which was misinterpreted as an extracellular loop in previous studies (*8*, *10*, *20*–*22*). The end of TM2 turns 180° in the membrane and connects to TM3, which re-enters the cytoplasm. A tryptophan (W88) at the junction between TM2 and TM3 is conserved among cornichons (fig. S15). The folded jackknife shape of TM2 and TM3 may be a signature of all cornichons. The majority of CNIH3 is embedded in the membrane with a small cytoplasmic domain, and thus a direct interaction between the LBD of GluA2 and CNIH3 in the extracellular space is unlikely. In contrast, TARPs and GSG1L modulate AMPARs by directly contacting the LBD in the extracellular space (*24*, *25*). Based on the structure and the locations of functionally important mutations of CNIH3, receptor modulation must occur via the intramembrane and the cytoplasmic interaction between the two proteins (*21*, *26*) (see supplementary text).

CNIH3 binds to the M1 and M4 of adjacent subunits of GluA2, where TARPs and GSG1L associate (Fig. 2D) (*15*, *16*). Despite the absence of homology, the bundle of four helices of CNIH3 resembles the geometries of those of TARPs and GSG1L. Geometric conservation extends to the M1 and M4 of GluA2, which interface with the auxiliary subunits. Indeed, the helices of CNIH3 and TARP γ-8 together with the M1 and M4 of GluA2 can be superimposed, when the Cα backbones of M4 of both complexes are

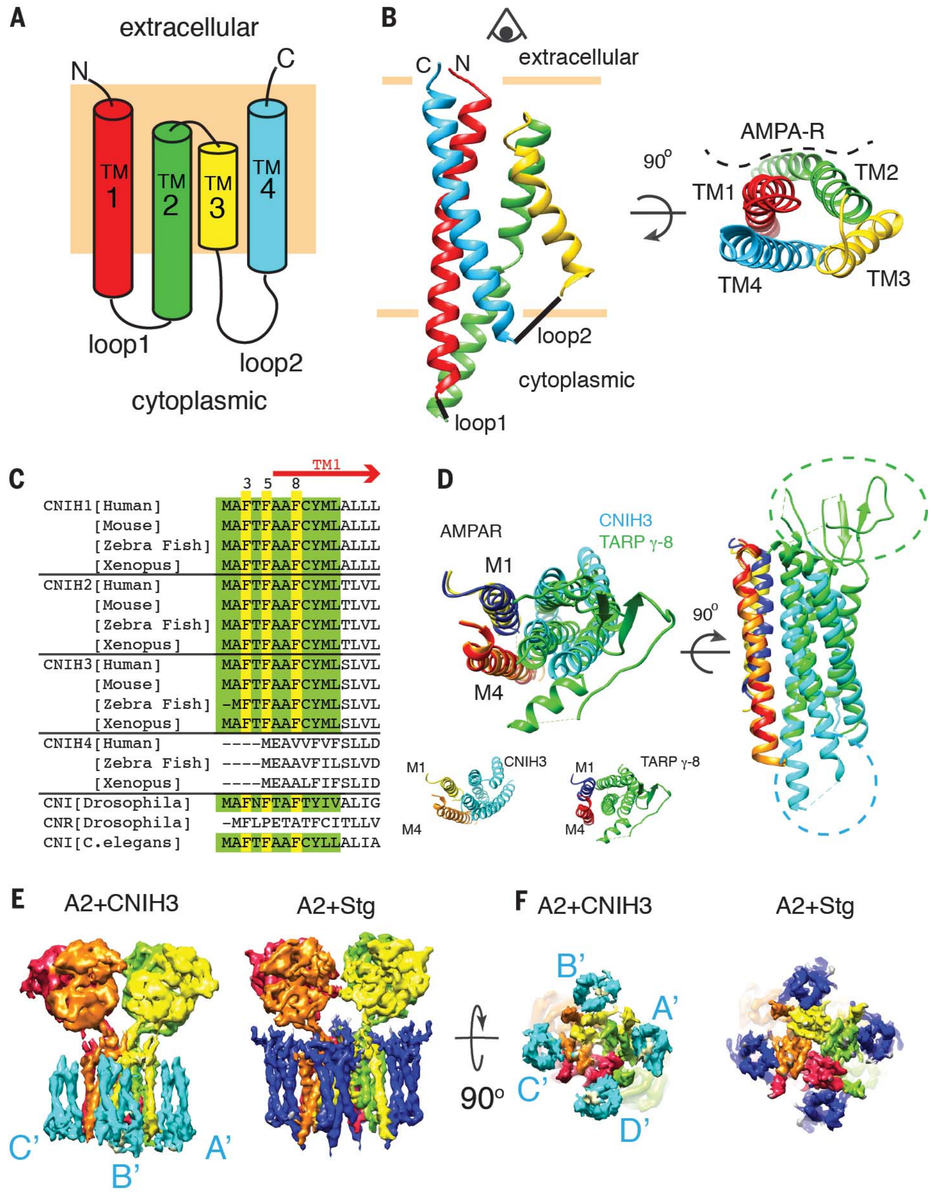


Fig. 2. Membrane topology of CNIH3. (**A**) Schematic of the topology. The rectangle (pale orange) in the background represents the membrane. (**B**) Ribbon diagram of CNIH3, from side (left) and top (right) views. The helices are colored according to the topology diagram in (A). The location of AMPAR is shown in the top view. The model was built from map LBD-TMD-C3 (table S1). (**C**) Sequence alignment of the N-terminal fragment that contains the UMIP (green and yellow). F3, -5, and -8 (yellow) play critical roles in complex assembly (see Fig. 3E). (**D**) The M1 and M4 of adjacent subunits of AMPAR form the binding surface for both CNIH3 and TARP. Aligning the M4 helix is sufficient to superimpose the remaining helices (M1 and TM1 to TM4 of CNIH3 or TARP γ-8). The models before alignment are shown on the bottom. The side view reveals the distinctive extracellular and cytoplasmic extensions of TARP γ-8 (green) and CNIH3 (cyan). (**E** and **F**) Density maps of the GluA2-CNIH3 complex in PS [map A2-C3(PS) at 6.91σ] compared with that of the GluA2-stargazin complex (EMDB-8721 at 25.1σ) at a comparable resolution. The NTD is excluded from the display. CNIH3 and stargazin are shown in cyan and blue, respectively. Side and bottom views are shown.

aligned at an RMSD of 0.763 Å (Fig. 2D). The detailed residue contacts are different (see the supplementary text and fig. S16), even though the overall architectures of the complexes appear similar at low resolution (Fig. 2, E and F).

Numerous nonprotein densities (L-a to L-h), whose features are characteristic of either lipids or detergents, surround the hydrophobic surface (Fig. 3, A and B). These densities were best resolved in the map LBD-TMD-C3$_{\text{lipid}}$, which was calculated from a smaller number of particles than that used to generate the higher-resolution map LBD-TMD-C3 (see materials and methods and fig. S9), indicative of a small structural (or occupancy) variability in the lipids and detergents that are associated with the complex. L-a to L-e and L-h were visible at a wide range of thresholds and modeled as acyl groups. L-h is a component of the inner (cytoplasmic) leaflet of the lipid bilayer and occupies the space where M2 is typically located, making contacts with both GluA2 and CNIH3 (fig. S17). The tips of L-c and L-h contact each other near the center of the membrane (Fig. 3D). L-c, an outer leaflet component, contacts both M1 (Y523 and V530) and M3 (F607) of adjacent subunits (Fig. 3C and fig. S17), and it approaches the center of the ion channel, similar to the lipid density L4 found in the heteromeric AMPAR in complex with TARP γ-8 (*25*). However, the contact points of L-c and L4 with AMPAR are different, except for F607 of M3. I hypothesize that the contrasting lipid geometries of AMPAR in complex with TARP γ-8 and CNIH3 contribute to their functional differences.

A bulkier lipid (L-f), specific to A2-C3 and interpreted as a cholesterol group, sits next to the interaction interface between GluA2 and CNIH3, making contacts with both M4 (Y797) and TM1 (L157 and M153) (Fig. 3D and fig. S17). Within the interface, three phenylalanines (F3, -5, and -8) of CNIH3 make contacts with M1 (E524, M527, and C528) and M4 (L789, A793, and Y797) of GluA2 (Fig. 3E). In particular, Y797 interacts with both L-f and CNIH3. Except for Y797, the residues at the interface are specific to AMPARs and are replaced by different residues in closely related kainate receptors that do not interact with CNIH3, establishing specificity for assembly. A previous study demonstrated that introducing mutations to residues L528, L789, and A793, now shown to be at the interface, destabilizes

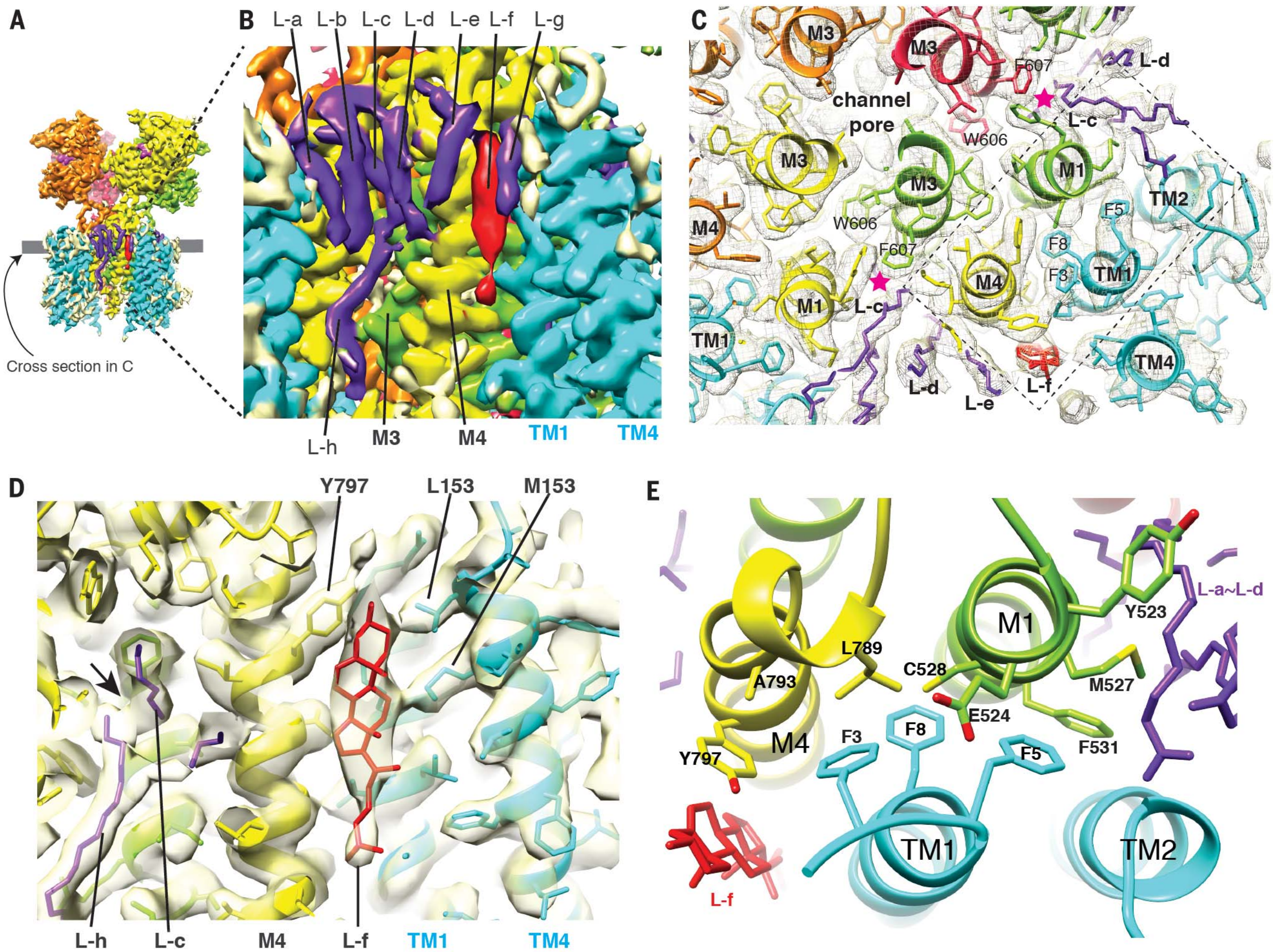


Fig. 3. Binding interface and arrangements of lipids. (**A**) The densities that surround the transmembrane helices have a characteristic appearance of densities derived from lipid and detergents. The map LBD-TMD-C3$_{\text{lipid}}$ (see materials and methods and fig. S9) is displayed at 6.02σ. (**B**) Magnified view of the area in (A) where the lipid-like densities are attached to the complex (L-a to L-h). The tips of L-c and L-h make contact. L-f is a bulkier density whose contour resembles a cholesterol group. The visualization threshold was set at 6.02σ. (**C**) Cross section at the level indicated by the gray line in (A). The map and model are superimposed. Stars (magenta) indicate contacts between L-c and F607(M3). The details within the dashed rectangle are shown in (E). (**D**) Superimposed density map and a molecular model, showing contacts made between L-f (cholesterol) and side chains of GluA2 and CNIH3. The tips of L-c and L-h make contact at the arrow. The visualization threshold was set at 5.32σ. (**E**) Molecular architecture of the GluA2/CNIH3 interface. F3, -5, and -8 of CNIH3 (cyan) contact residues in M1 (E524, M527, C528, and F531) and M4 (L789, A793, and Y797) of adjacent subunits of GluA2. Y797 (M4) and M527 (M1) simultaneously contact lipid-like densities L-f and L-b, respectively. Y523 (M1), which does not contact CNIH3, is immediately next to the interface and contacts L-c. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

the complex and bidirectionally alters the magnitude of gating modulation by CNIH3 (*26*). Replacing A793 of M4 and F3 of CNIH3 with cysteine induced a disulfide cross-link, further supporting the model that two residues interact (fig. S18). I hypothesize that the binding site for the three phenylalanines (F3, -5, and -8 of CNIH3) near the extracellular surface of GluA2 dictates gating modulation and is a potential target for drugs that could be used to control the ion channel activity of AMPARs. The detailed architecture of the interaction interface suggests a possible role for lipids in regulating the assembly or function of the complexes formed by AMPAR and auxiliary subunits. The molecular model of CNIH3 is likely to serve as a reference for future investigations of the biology of the cornichon family.

REFERENCES AND NOTES

1. S. F. Traynelis *et al.*, *Pharmacol. Rev.* **62**, 405–496 (2010).
2. D. Bowie, *CNS Neurol. Disord. Drug Targets* **7**, 129–143 (2008).
3. T. Nakagawa, Y. Cheng, E. Ramm, M. Sheng, T. Walz, *Nature* **433**, 545–549 (2005).
4. N. Armstrong, E. Gouaux, *Neuron* **28**, 165–181 (2000).
5. A. I. Sobolevsky, M. P. Rosconi, E. Gouaux, *Nature* **462**, 745–756 (2009).
6. B. Herguedas *et al.*, *Science* **352**, aad3873 (2016).
7. A. C. Jackson, R. A. Nicoll, *Neuron* **70**, 178–199 (2011).
8. J. Schwenk *et al.*, *Neuron* **74**, 621–633 (2012).
9. N. F. Shanks *et al.*, *Cell Rep.* **1**, 590–598 (2012).
10. J. Schwenk *et al.*, *Science* **323**, 1313–1319 (2009).
11. B. E. Herring *et al.*, *Neuron* **77**, 1083–1096 (2013).
12. P. J. Brockie *et al.*, *Neuron* **80**, 129–142 (2013).
13. X. Gu *et al.*, *Nat. Commun.* **7**, 10873 (2016).
14. S. Boudkkazi, A. Brechet, J. Schwenk, B. Fakler, *Neuron* **82**, 848–858 (2014).
15. E. C. Twomey, M. V. Yelshanskaya, R. A. Grassucci, J. Frank, A. I. Sobolevsky, *Nature* **549**, 60–65 (2017).
16. S. Chen *et al.*, *Cell* **170**, 1234–1246.e14 (2017).
17. M. V. Yelshanskaya, M. Li, A. I. Sobolevsky, *Science* **345**, 1070–1074 (2014).
18. N. Tajima *et al.*, *Nature* **534**, 63–68 (2016).
19. Y. Zhao, S. Chen, C. Yoshioka, I. Baconguis, E. Gouaux, *Nature* **536**, 108–111 (2016).
20. M. M. Wudick *et al.*, *Science* **360**, 533–536 (2018).
21. N. F. Shanks *et al.*, *J. Neurosci.* **34**, 12104–12120 (2014).

22. Y. Shi *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16315–16319 (2010).
23. J. Traynelis *et al*., *Genome Res.* **27**, 1715–1729 (2017).
24. E. C. Twomey, M. V. Yelshanskaya, R. A. Grassucci, J. Frank, A. I. Sobolevsky, *Neuron* **94**, 569–580.e5 (2017).
25. B. Herguedas *et al*., *Science* **364**, eaav9011 (2019).
26. N. M. Hawken, E. I. Zaika, T. Nakagawa, *J. Physiol.* **595**, 6517–6539 (2017).

ACKNOWLEDGMENTS

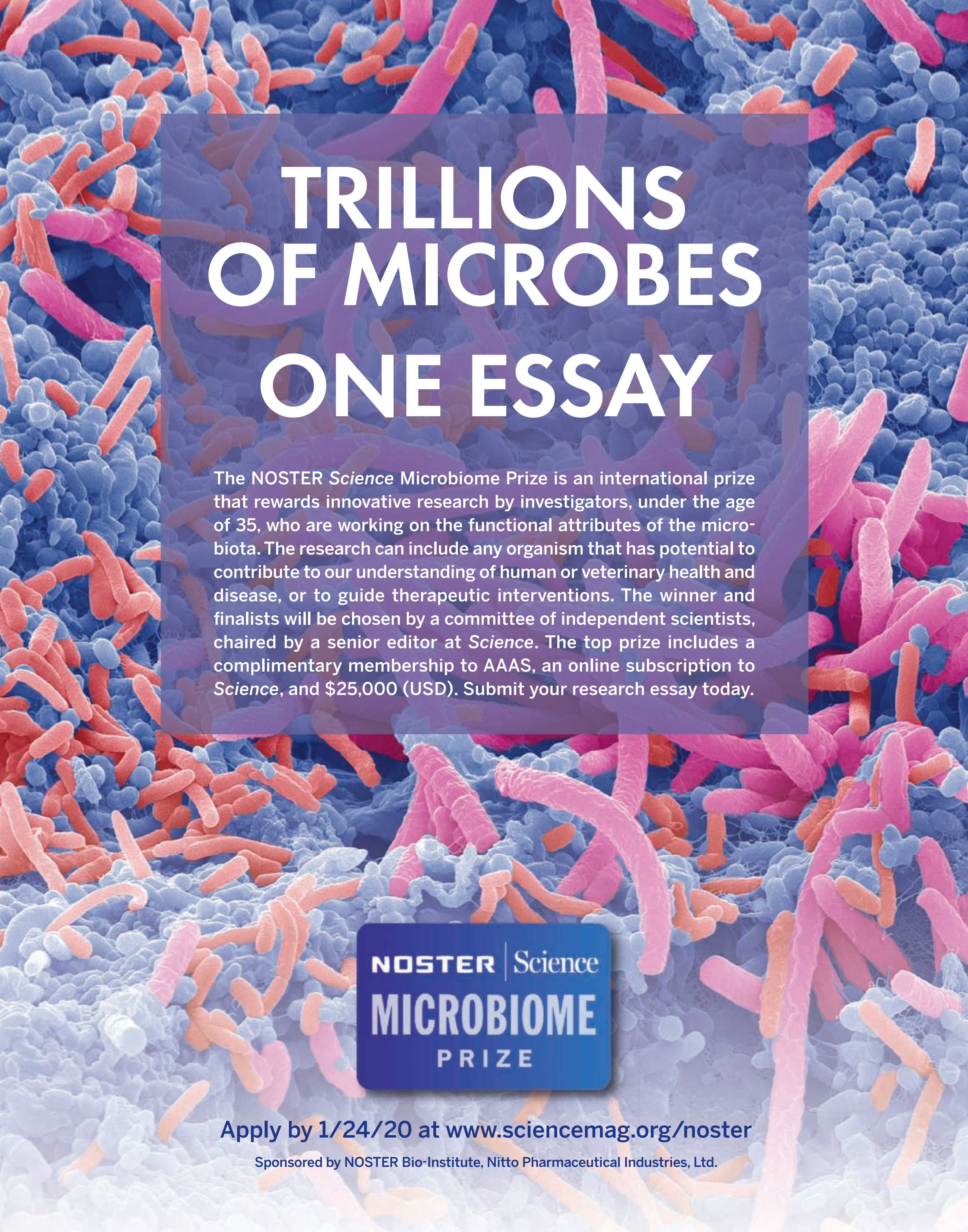
I thank E. Zaika for technical support in isolating clone #7. I acknowledge the use of the molecular cryo-EM facility (supported by S. Collier, E. Binshtein, and M. Chambers) and computation resources (ACCRE and DORS) at Vanderbilt University. I thank C. Azumaya, A. Kamalova, E. Karakas, C. Sanders, R. Danev, and I. Greger for discussions. I thank M. Kikkawa and H. Yanagisawa for providing access and facilitating data collection using the Titan Krios instrument at the University of Tokyo. I thank W. Chazin and R. Colbran for support. Software was provided by SBgrid. **Funding:** This work was supported by funding from NIH (R01HD061543 to T.N.) and Vanderbilt University (to T.N.). DORS was supported by NIH (S10RR031634 to J. Smith). **Author contributions:** T.N. conceived and designed the project, conducted experiments, analyzed results, interpreted data, and wrote the manuscript. **Competing interests:** The author declares no competing interests. **Data and materials availability:** Reagents and other materials will be available upon request to T.N. The structural data have been deposited in the Electron Microscopy Data Bank (EMDB) with IDs 20330, 20332, 20645, 20666, 20717, 20727, 20732, 20733, and 20734 and in the Protein Data Bank (PDB) with IDs 6PEQ, 6U5S, 6U6I, 6UCB, 6UD4, and 6UD8.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6470/1259/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Table S1
References (*27–40*)

View/request a protocol for this paper from *Bio-protocol*.

5 June 2019; accepted 21 October 2019
10.1126/science.aay2783




NEVER STOP
ACCELERATING REGENERATIVE MEDICINE
We're applying our photographic film innovations to help advance new treatments in the revolutionary field of regenerative medicine. Over the last 80-plus years, we've developed advanced technology that controls complex chemical reactions in photographic film that's a mere 20 microns(*1) thick. And today, that technologyis being applied to research and the world's first clinical trial(*2) of medical treatments that use high-quality iPS cells. And in the future, we'll strive to help those suffering from a range of medical conditions, such as those of the eyes, nerves, heart and more. Of course, the challenges are endless, but so are the possibilities. Which is why we'll never stop accelerating regenerative medicine to help build astronger, healthier future for all.
FUJIFILM
Value from Innovation
Follow Fujifilm Life Sciences at in
*1 Thickness of layers excluding the base.
*2 Fujifilm's iPS cells are being utilized in the world's first clinical trial using iPS cells conducted in the UK by the Australian company Cynata.
FUJIFILM and Fujifilm Value from Innovation are trademarks of FUJIFILM Corporation.
©2019 FUJIFILM Corporation. All rights reserved.
FUJIFILM

Exome Capture Kit
ExomeCG is a clinically enhanced exome capture kit that—for the first time—allows confident, robust whole-exome sequencing and targeted copy number analysis in a single assay. The clinically validated test replaces the need for chromosomal microarray and multiplex ligation-dependent probe amplification (MLPA) front line tests, saving time and money while achieving the highest diagnostic yield possible. ExomeCG is designed to provide unparalleled coverage of clinical targets when used in combination with the Congenica clinical decision support platform, which enables fast, accurate interpretation of next-generation sequencing data for health care professionals, helping them to deliver world-class genomic medicine services and make important clinical decisions.
Nonacus
For info: +44-(0)-7811-996-942
www.nonacus.com

Touchscreen Motorized Repeating Pipette
The BRAND HandyStep touch motorized repeating pipette features a touchscreen interface for intuitive operation. The large color display provides all pertinent information and allows easy adjustment of all parameters. It can be used with all standard repeating pipette tips. Automatic size recognition of BRAND PD-Tip II precision dispenser tips adds to the ease of use and reduces the chance of errors. Two models are available: The HandyStep touch features dispensing, auto-dispensing, and single-volume pipetting modes, with the ability to save up to 10 favorites. The auto-dispensing mode features a “learn” function for easy interval programming. The HandyStep touch S adds multiaspiration, sequential dispensing, and titration modes.
BrandTech Scientific
For info: 888-522-2726
www.brandtech.com/product/new-handystep-touch

mRNA Transfection Reagent
Polyplus-transfection announces the launch of in vivo-jetRNA, an injectable messenger RNA (mRNA) transfection reagent. Additionally, its adaptability enables multiple administration routes to target an organ of choice in all types of animals—especially the spleen and lymphatic nodes that play a major role in vaccination. As plasmid DNA, mRNA can be used to transiently express a specific antigen into an organism or tissue. This antigen will be recognized by the immune system through antigen-presenting cells and triggers a specific immune response. It is also beneficial for cancer research, as it can be used to express a suicide gene to kill cancer cells.
Polyplus-transfection
For info: +33-(0)-3-90-40-61-80
www.polyplus-transfection.com/products/invivo-jetrna

Knockout Lysates
Abcam’s new engineered knockout (KO) cell lysates enable proteomic studies by providing researchers with true negative controls. Acquiring reliable, off-the-shelf KO mouse models or cell lines that match specific experimental requirements can be challenging; these often need to be sourced directly from individual researchers or labs. Abcam’s newly launched collection addresses this issue by providing access to over 2,800 diploid KO cell lysates. These lysates are useful for studies requiring loss-of-function phenotypes at the proteomic level as well as routine applications such as Western blotting and mass spectroscopy. They can also be used to support antibody and target validation. The use of diploid cells makes them well suited to the more complex studies often carried out in cancer research. Derived from commonly used immortalized cell lines, KO lysates have been engineered using CRISPR-Cas9 and are accompanied by Sanger sequencing and Western blotting validation data. Corresponding wild-type controls are also provided so that the biological impact of each KO lysate can easily be assessed within a consistent cellular background.
Abcam
For info: 888-772-2226
www.abcam.com

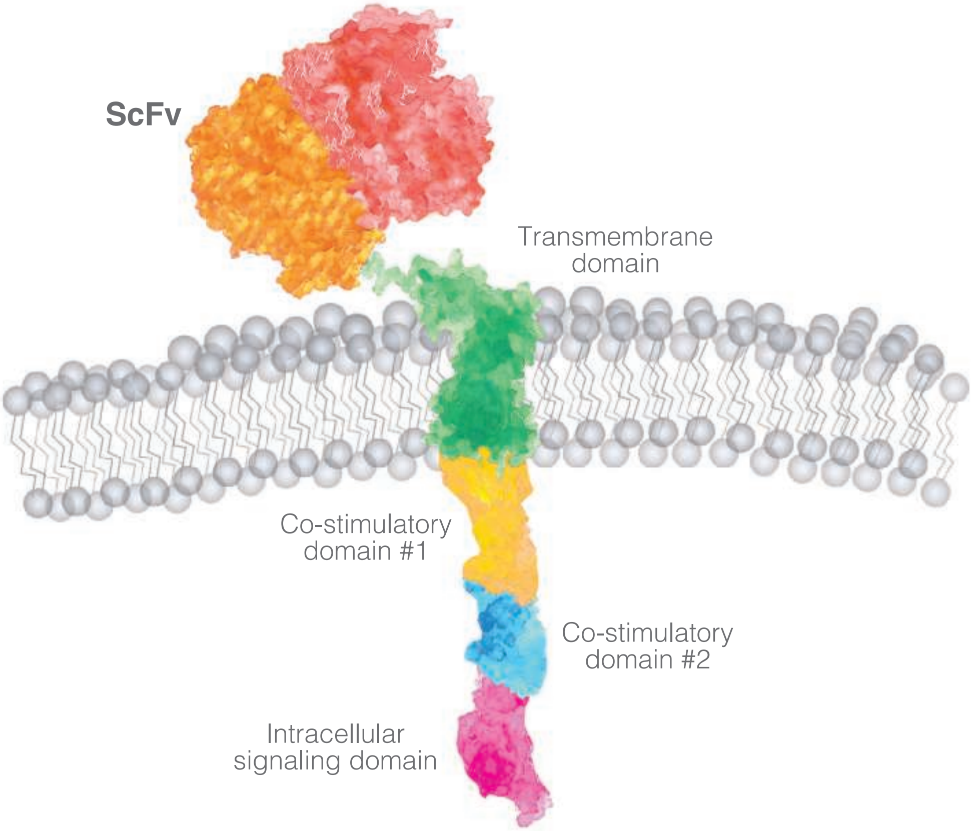
Protein Extraction Kit
The Chromatrap Protein Extraction kit from Porvair Sciences offers a fast, efficient method of extracting cytoplasmic, nuclear, or total protein from mammalian cells. Combining the optimized protocol with proprietary reagents, the kit provides high yields of purified active protein with low cross-fraction contamination in less than 90 min. It is easy to use, avoiding the need for ultracentrifugation steps and delivering reproducible preparation of purified protein extracts from a wide variety of biological samples. Affordably priced, the kit generates high-quality protein samples that may be directly applied to several applications, including Western blotting, gel-shift assays, protein assays, reporter gene assays, and enzyme activity assays.
Porvair Sciences
For info: +44-(0)-1978-661144
www.chromatrap.com/protein-extraction-kit

Automated Perfusion Cell Culture System
The Corning MicroDEN is an automated perfusion cell culture system that provides a more efficient process for the differentiation of monocytes into dendritic cells than current manual methods. It enables the wide use of dendritic cells in cell and gene therapy for basic and applied research in immunology and immuno-oncology. Developed by Flaskworks, a Boston-based startup company specializing in technologies for therapeutic cell production, MicroDEN is now available from Corning. The rise of personalized immunotherapies has increased demand for dendritic cells. However, the current process for generating these cells is cumbersome, error prone, and time intensive. Dendritic cells are key elements in immunology and immunotherapy research in oncology and infectious diseases as well as autoimmune diseases and transplant rejection. MicroDEN streamlines the process by minimizing manual steps, resulting in more consistent yields, reduced risk of contamination, and more efficient therapy research.
Corning
For info: 800-492-1110
www.corning.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

CAR-T/NK Cells

Ready-to-use CAR-T products to advance your research



Current CAR-T Products

• Her2	• CD22	• CD47	• CD19	• Mesothelin
• Her2-GFP	• CD33	• CS1	• CD19-FLAG	• Mesothelin-FLAG
• EGFR	• CD38	• BCMA	• CD19-GFP	• EpCAM
• CD19-CD22	• iCas9	• GPC3	• CD4	• ROR-1
• CEA	• CD24	• CD37	• PDL-1	• B7H4

And many more. Custom development services available.

ProMab Biotechnologies has developed over 60 types of off-the-shelf CAR-T/NK cells, target cells and generated CAR-T with several patented innovations in the CAR construct design. ProMab can initiate your CAR project in many ways: starting from your antibody sequence, by developing a mouse monoclonal or humanized antibody; or even producing a fully human single-chain variable antibody fragment (scFv) from our existing antibody library. ProMab Biotechnologies has completed more than 300 CAR-T projects and has developed about 5,000 hybridoma cell lines to date. Inquire today about your custom project.

All products are for research only

USDA Agricultural Research Service

www.ars.usda.gov

The USDA, Agricultural Research Service, Grand Forks Human Nutrition Research Center in Grand Forks, ND, is seeking a full time **POSTDOCTORAL RESEARCH ASSOCIATE**, (Research Chemist) for a TWO-YEAR APPOINTMENT. Ph.D. is required. Salary is commensurate with experience ($62,236- $80,912per annum) plus benefits. Citizenship restrictions apply.

The incumbent will develop and implement mass spectrometry-based analytical methods for lipidomic characterization of dairy products. Knowledge of the principles, theories and practices of chemistry, analytical chemistry and a working knowledge of analytical instrumentation including mass spectrometers are desirable. Send application materials including curriculum vitae, transcripts, and references to:

Dr. Matthew Picklo,
Grand Forks Human Nutrition Research Center
2420 2nd Ave N
Grand Forks, ND 58203
or
e-mail matthew.picklo@usda.gov

USDA/ARS is an Equal Opportunity Provider and Employer.

The University of Vermont
LARNER COLLEGE OF MEDICINE
Department of Pharmacology

Pharmacology Tenure Pathway Faculty Position

The Department of Pharmacology at the University of Vermont invites applications for a faculty position at the Assistant Professor level on the Tenure Pathway, although Associate and Full Professor candidates will be considered.

Applicants should have a Ph.D. and/or M.D. degree and be committed to excellence in both research and the teaching of Pharmacology. We are seeking candidates who have a record of research excellence in areas that complement existing programs within the Department and who are applying innovative research strategies and approaches. Our faculty have diverse interests that include neurovascular and cardiovascular regulation and cellular signaling mechanisms associated with brain health and disease states. The candidate is expected to maintain an independent, extramurally funded research program. The candidate also must be willing to team-teach pharmacology in undergraduate, medical, and graduate school courses. Start-up funds will be competitive.

The University of Vermont is especially interested in candidates who can contribute to the diversity and inclusive excellence of the academic community through their teaching, service and research, scholarship or creative arts. Applicants are required to submit a separate statement of advancing diversity and inclusive excellence.

We are an educationally purposeful community seeking to prepare students to be accountable leaders in a diverse and changing world. Members of the University of Vermont community embrace and advance the values of Our Common Ground: Openness, Respect, Responsibility, Integrity, Innovation, and Justice. The successful candidate will demonstrate a strong commitment to the ideals of accessibility, inclusiveness, and academic excellence as reflected in the tenets of Our Common Ground.

The University of Vermont is an Equal Opportunity/Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, protected veteran status, or any other category legally protected by federal or state law. The University encourages applications from all individuals who will contribute to the diversity and excellence of the institution.

Review of applications will begin immediately and continue until the position is filled. Include a CV, research plan, teaching statement and experience, diversity statement, and the names and email addresses of three references. After the initial review of applications, references may be contacted to submit their letters directly to the search committee.

Interested candidates must apply online at: **https://www.uvmjobs.com/postings/38276** under Position No. 00025109. Inquiries about the position should be addressed to: Dr. Joseph Brayden, Univ. Vermont, Dept. of Pharmacology, Given Medical Building, 89 Beaumont Avenue, Burlington, VT 05405-0068 USA, or via email at: joe.brayden@uvm.edu.

SANFORD HEALTH

Assistant Scientist/Professor in Biomedical Research, Sanford Research

Sanford Research invites applications for full-time faculty at the rank of Assistant Scientist within Sanford Research in Sioux Falls, SD, with commensurate rank of Assistant Professor at the Sanford School of Medicine at the University of South Dakota. Sanford Research is the non-profit reasearch branch under Sanford Health. We seek outstanding scientists with research programs that span areas of biomedical research addressing pediatric and/or rare diseases. This includes, but is not limited to, investigating developmental, cellular and molecular disease mechanisms, using animal models to understand disease progression and design therapeutic interventions, and developing new therapeutic approaches. Successful candidates will join a collaborative team with expertise spanning genetics and genomics, stem cell biology, cell signaling, developmental biology, and cancer biology as they pertain to pediatric and/or rare diseases. Significant institutional support, including modern laboratory space and state-of-the-art facilities will be provided. A comprehensive benefits package will be tailored to the candidate's qualifications.

Qualifications: Applicants should hold a PhD, MD or MD/PhD degree and complement the existing strengths and interdisciplinary and collaborative nature of Sanford Research. Physician Scientists are encouraged to apply. Candidates will be expected to develop independent research programs and secure extramural funding.

Application: Sanford Health is an Equal Opportunity/Affirmative Action Employer. Applicants should submit a single PDF that includes: (1) detailed curriculum vitae, (2) description of research experience and future research plans with details on relevance of their research to genetics and genomics, pediatrics and rare diseases, stem cell biology, diabetes, or environmental influences of diseases, and (3) three letters of recommendation. If any of the information above is missing, the submission will not be considered. Submit materials via email to: **researchrecruitment@sanfordhealth.org**

About Sanford Health: Sanford Health, one of the largest health systems in the United States, is dedicated to the integrated delivery of health care, genomic medicine, senior care and services, global clinics, research and affordable insurance. Headquartered in Sioux Falls, South Dakota, the organization includes 44 hospitals, 1,400 physicians and more than 200 Good Samaritan Society senior care locations in 26 states and nine countries. Nearly $1 billion in gifts from philanthropist Denny Sanford have transformed how Sanford Health improves the human condition.

UNIVERSITY OF GEORGIA

College of Public Health
Department of Environmental Health Science

A Tenure-Track Assistant Professor of Environmental Health Science

The Department of Environmental Health Science in the College of Public Health at the University of Georgia invites applications for a tenure-track Assistant Professor. The position is an academic year (9 month) appointment with an opportunity to supplement salary through external funding. We are seeking applicants with a strong academic record in teaching, research, and service in the area of environmental health sciences with focus on studying adverse health effects of environmental hazards on human populations. Qualifications should include: doctoral degree in a discipline related to environmental and occupational health, environmental epidemiology, biochemistry, or molecular toxicology with at least two years of postdoctoral training. Preferred qualifications include evidence of the ability to submit strong (fundable) proposals for extramural funding, ability to produce and communicate research, and excellence in teaching and working with students. A commitment to cultural diversity is expected. The successful candidate will be required to teach at least one undergraduate and one graduate course in addition to mentoring student research.

Review of applications will begin **November 15th, 2019** and will continue until the position is filled. Application materials should include a cover letter, current curriculum vitae, statement of research and teaching experience, and the names and contact information of three references (name, address, e-mail address, and phone number). Materials should be submitted via the UGA Job Search site: **http://www.ugajobsearch.com/postings/125630**.

The University of Georgia is an Equal Opportunity/Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, ethnicity, age, genetic information, disability, gender identity, sexual orientation or protected veteran status. Persons needing accommodations or assistance with the accessibility of materials related to this search are encouraged to contact Central HR (hrweb@uga.edu). Please do not contact the department or search committee with such requests.

206 Environmental Health Science Building ● Athens, Georgia 30602-2102
(706) 542-2454 ● FAX (706) 542-7472

ETH zürich

Assistant Professor of Molecular and Structural Biology

→ The Department of Biology (www.biol.ethz.ch) at the ETH Zurich invites applications for an Assistant Professorship in Molecular and Structural Biology at the Institute of Molecular Biology and Biophysics (www.mol.biol.ethz.ch).

→ We are looking for researchers who investigate the structure and mechanism of biological macromolecules, molecular assemblies or cellular machineries using diverse structural, biochemical, biophysical or computational techniques. Successful candidates are expected to establish and lead a competitive research team, commit to teaching at the Department of Biology, and to contribute to the interactive, scientific environment at ETH Zurich.

Assistant professorships have been established to promote the careers of younger scientists. The initial appointment is for four years with the possibility of renewal for a three-year period.

→ **Please apply online: www.facultyaffairs.ethz.ch**

→ Applications should include a curriculum vitae, a list of publications, a statement of future research and teaching interests, the names of at least three references, and a description of the three most important achievements. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Joël Mesot. The closing date for applications is 1 February 2020.** ETH Zurich is an equal opportunity and family friendly employer, strives to increase the number of women professors, and is responsive to the needs of dual career couples.

FOCUS ON DIVERSITY AND INCLUSION

UNIVERSITY OF GEORGIA

1785

Assistant/Associate or Full Professor in Physiology

The Department of Physiology and Pharmacology at the University of Georgia, College of Veterinary Medicine, invites applications for a tenure-track faculty position at the rank of Assistant, Associate, or Full Professor.

We seek outstanding candidates with active research programs applying innovative multidisciplinary approaches, including models of disease and development to the study of physiology. Candidates with research interests in cardiovascular, respiratory or gastrointestinal physiology as well as models of animal/human disease and development are encouraged to apply. The successful candidate will be expected to establish a robust externally funded research program and contribute to the Department's teaching mission in his/her area of expertise. Applicants must have a PhD, DVM/PhD or MD/PhD degree and post-doctoral experience in Physiology or a related field as well as a publication record appropriate for the rank. Assistant Professor level candidates must have at least two years of relevant postdoctoral research and present a well-developed research plan. Associate or Full Professor level applicants should show evidence of an established, well-funded, and impactful research program. Candidates should also have a commitment to teaching excellence and one semester of prior teaching experience. Applicants should submit a letter of intent, curriculum vitae, research statement, teaching statement and contact information for at least three references to the following website: **http://www.ugajobsearch.com/postings/128983**. Applications received by **January 10, 2020** are assured full consideration. An attractive start-up package with competitive conditions of employment will be offered. The anticipated start date is August 1, 2020. Persons needing accommodations or assistance with the accessibility of materials related to this search are encouraged to contact Central HR (hrweb@uga.edu).

The University of Georgia (UGA), the nations' first chartered state-supported university, is a research-intensive land-grant and sea-grant university. The University of Georgia is an Equal Opportunity/Affirmative Action employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, ethnicity, age, genetic information, disability, gender identity, sexual orientation or protected veteran status.

FOCUS ON DIVERSITY AND INCLUSION

Massachusetts Institute of Technology

Department of Biological Engineering
Division of Comparative Medicine
Director/Tenured Faculty Position

The MIT Division of Comparative Medicine in partnership with the MIT Department of Biological Engineering, invites applications for the Director, at the tenured faculty level, with a start date of 1 July 2020 or a mutually agreeable date thereafter. Applicants should hold a DVM from an AVMA accredited veterinary school by the beginning of employment, with demonstrated expertise in a discipline which applies molecular/cellular bioscience to the study of the microbial/host interface in health and disease with interests in microbial pathogenesis, microbiome bioinformatics, immunology, oncogenesis, or pathobiology. The selected candidate will develop and sustain a vigorous extramurally funded research program, supervise graduate and postdoctoral fellows and teach graduate level classes.

The Division of Comparative Medicine has three basic missions: education, research, and the provision of managing and overseeing a comprehensive animal husbandry, clinical, and diagnostic program for all research animals maintained at MIT. The Director will provide direction and administrative coordination to facilitate the Division's role in providing expertise and support to investigators using animals in biomedical research and teaching; to conduct grant funded research and develop animal models pertinent to biomedical research; to provide postdoctoral training to veterinarians seeking careers in biomedical research; develop scientists within the Division who can successfully compete for grants.

The Division, consisting of approximately 180 personnel, has centralized administrative responsibility for all AAALAC accredited facilities, consisting of 193,000 square feet on the MIT campus. This responsibility underlies the overall commitment of the Institute in providing a comprehensive animal care and use program.

Candidates must register with the BE search website at http://be-fac-search.mit.edu, and must submit application materials electronically to this website. Candidate applications should include a description of professional interests and goals in both teaching and research. Each application should include a curriculum vitae and the names and addresses of three or more references who will provide recommendation letters. References should submit their letters directly at http://be-fac-search.mit.edu.

Applications received by December 15, 2019 will be given priority.

Questions may be directed to: Prof. Angela Belcher (bedh@mit.edu), Head - Department of Biological Engineering, MIT 16-343 Cambridge MA 02139.

MIT is an Equal Opportunity/ Affirmative Action employer

http://web.mit.edu

Focus on China

CHINA 聚焦“中国”

Zhimin Li
Director of Center for Science and Technology Development Ministry of Education, People's Republic of China

As its civilization evolved over the centuries, China developed a well-organized knowledge base for agriculture, medicine, astronomy, and mathematics, and powerfully influenced our world with great scientific contributions, including the four great inventions (i.e., papermaking, gunpowder, printing, and the compass), and many other world-renowned technical achievements. Since its foundation in 1949, the People's Republic of China has built a system of research institutes as well as an echelon of top-notch researchers. Policies have been implemented providing full aid, comprehensive management, and support services for scientific research; and as a result, hundreds of institutes, universities, and R&D departments in companies across China are making breakthroughs that have influenced scientists around the world.

Recent years have seen the rapid growth of research funding in China, which totaled RMB 1.96 billion (USD 280 million) in 2018. The over 20% average annual increase in research investment during the past 20 years is far above the country's GDP growth for the same period. China has been the world's second-largest-spending country in terms of research, providing strong financial support for its scientific development. Its scientific competence has consistently improved, and China has equaled or surpassed other nations in some frontier scientific areas. A large number of significant discoveries have laid the foundation for China to become a scientific and technological powerhouse. China now stands in a transformational period as it moves from focusing merely on quantity to emphasizing quality, and from isolated breakthroughs to a systematic upgrade in scientific research.

Throughout the history of human society, scientific innovation has always been a driving force for the development of a nation, and for the progress of civilization as a whole. The Chinese government has a historic opportunity to respond to the call to be at the forefront of science and technology—the most crucial sector of the economy—and to meet the most pressing needs of the country, accelerate innovation in key fields, promote the acquisition of new talents, and take the lead in global scientific competition. In order to achieve the goal of modernization, China will focus on the following aspects:

Improve the fundamental capacity of science and technology.

Several national laboratories have been established to explore the critical scientific questions of our time, such as the structure of matter, the origin of the universe, the evolution of life on earth, and the nature of consciousness, and to expand our knowledge of cutting-edge fields such as information technology, artificial intelligence, and robotics. At this time it is essential to implement the construction of national laboratories that focus on the major scientific and technological tasks at hand, and to develop large-scale infrastructure that relies on the most advantageous and innovative systems and comprehensively integrates available resources. All of this is necessary in order to establish a new mechanism of target-orientation, performance management, collaborative research, and open sharing, and to ensure that national laboratories are characterized by advancement, expert guidance, and platform support.

Increase access to technology.

Science and technology should be able to address the major issues people face in their daily lives—issues such as disease prevention and control, food and drug safety, and population aging. We

should work to increase the availability of technology to the public, and contribute to building a low-cost, wide-coverage, high-quality public service system that helps people enjoy a livable environment, delivers better medical and health services, and assures them of adequate food.

Put technology at the service of economic and social development.

Scientific and technological innovation should enable us to promote sustained and healthy economic and social development by allowing us to implement major national science and technology programs; resolve national targets; and further highlight key areas of growth; and to assign major strategic products, key common technologies, and large-scale projects as major special projects, thus effectively mobilizing national resources for great undertakings and, in turn boosting science and technology productivity.

Integrate science and technology innovation into the whole process of social development.

We are committed to create a culture of innovation: to let innovation become the starting point for social development—which is in itself the driving force behind innovation. We will improve the system of utilization, distribution, and income management of scientific and technological achievements, so that all institutions, talents, funds, and projects will be fully active and can jointly promote technological innovation. At the macro level, we will continue to initiate collaboration and to share new achievements in science and technology. We will participate in global governance in order to ensure common development within a community of broader interests. And we will also strive to create a favorable academic environment, one which encourages innovation and tolerates failure—while at the same time strengthening intellectual property protection.

Fully emphasize the cultivation of scientific and technological talents.

The key to technological innovation lies in talent; talent is at the core of all important technological developments. China has always attached great importance to cultivating talents. Recently, the central government issued Opinions on Deepening the Reform of the Talent Development System and Mechanism, to invigorate the country's human resource base. We must build a well-coordinated team of high-quality, innovative talents of every kind on a large scale, and we must develop their potential so as to mobilize the creative spirit that motivates outstanding scientific and technological personnel. If we gather first-class talent into the great cause of building our country, we will provide strong support for the realization of national development goals.

As the world enters a new era of new ideas and new challenges, it is inevitable that China should implement the strategy of rejuvenating itself through science and technology, talent cultivation, and innovation-driven growth. China must take these steps if it is to enhance its core competitiveness on the world stage, solve the deep-seated contradictions and problems involved in furthering its economic progress (e.g., how to deal with massive environmental changes), and build a prosperous society for all, while maintaining sustained and healthy economic and social development.

We welcome excellent scholars interested in applying for talent programs in China to contact us through AcaBridge(-consultant@acabridge.edu.cn), which provides one-on-one consultations.

SJTU Global Recruiting Program

Shanghai Jiao Tong University (SJTU) is one of the higher education institutions which enjoy a long history and a world-renowned reputation in China. Through some 120 years' unremitting efforts, SJTU has become a comprehensive, research-oriented, and internationalized top university in China.

SJTU now has 30 schools/departments, 31 research institutions, 13 affiliated hospitals, with around 50,000 students and over 3,000 full-time teachers, including the leading number of academic masters such as academicians of the Chinese Academy of Sciences and the Chinese Academy of Engineering, candidates for overseas talent programs and winners of National Outstanding Youth Fund among institutes of higher education in China.

Today SJTU has 67 undergraduate programs, 42 first-level disciplines authorized to offer doctorate degree, 57 first-level disciplines authorized to offer master degree. According to Thomson Reuter's Essential Science Indicator (ESI), SJTU has 19 disciplines listed World Top 1%, with 6 disciplines ranking World Top 0.1% and Engineering ranking World Top 0.01%. In 2019 QS World University Rankings by Subject, SJTU has 25 subjects ranking World Top 100, of which 10 subjects rank World Top 50. By the year of 2018, SJTU has led the country for the 9th consecutive year in terms of both the project number and the amount of money issued by National Natural Science Foundation of China. SJTU also ranks No.1 among all Chinese universities in the total number of published SCI papers and China excellent highly-cited papers.

Shanghai Jiao Tong University, carrying the mission of preserving cultural heritage, and seeking for the truth, bearing the responsibility of invigorating the Chinese nation and developing for the benefits of mankind, today this centennial university is sailing for the aim of becoming a comprehensive, research-oriented and internationalized world-class university. SJTU will provide free academic environment, strong research support and competitive compensation package for the talents. SJTU, your stage to becoming academic master!

Recruiting Position

Chair Professor/Distinguished Professor/Tenured Professor/Chief Researcher/Tenure-track Associate Professor

Work and Life Treatment

(1) Remuneration: With reference to the corresponding positions in the world's top universities, selected candidates will be provided with competitive remuneration and welfare benefits;
(2) Doctoral Students: A guaranteed number of doctoral students to be enrolled each year;
(3) Start-up Fund: Negotiable according to actual research demand;
(4) Housing: Assistance will be given in solving housing problem with furnished interim apartment provided;
(5) Healthcare: Health service will be provided depending on medical resources from 13 affiliated hospitals;
(6) Children's schooling: Children of pre-school and compulsory education age can be arranged to attend kindergartens and schools affiliated to SJTU.

Application Materials:

(1) Cover letter for the position that you are applying for;
(2) CV (with publication list);
(3) No less than 5 representative papers;
(4) Expertise and academic results.

How to Apply:

All the materials needed for application should be integrated into a PDF document named "name-position-department/school/discipline", to be sent to the email of Talent Resources Department (**ccy@sjtu.edu.cn**).

ShanghaiTech University is a young and dynamic higher education institution committed to carrying out China's national development strategy and nurturing the next generation of innovative scientists, inventors and entrepreneurs. With the backing and support of the Shanghai Municipal Government and China Academy of Science, ShanghaiTech's five schools, three research institutes and General Education Center seek cutting-edge solutions to address the challenges that China and the world is facing in the fields of energy, material, environment, human health, and artificial intelligence. As an integral part of the Zhangjiang Comprehensive National Science Center, ShanghaiTech is now leading several frontier research projects and large-scale facilities.

For more information, please visit: www.shanghaitech.edu.cn.

ShanghaiTech is now seeking talents in the following fields:

School of Physical Science and Technology: energy, system materials, photon and condensed state, material biology, environmental science and engineering

School of Life Science and Technology: molecular and cell biology, structural biology, neuroscience, immunology, stem cells and regenerative medicine, system biology and biological data, molecular imaging, biomedical engineering

School of Information Science and Technology: computer science, electrical engineering, information engineering, artificial intelligence, network and communication, virtual reality, statistics, big data and data mining

School of Entrepreneurship and Management: economics, finance, accounting, management, marketing, strategy and entrepreneurship

School of Creativity and Art: innovative design, filmmaking, game design, tech-driven art, big data visualization, creativity, design thinking

Shanghai Institute for Advanced Immunochemical Studies: antibody therapy, Immunotherapy, cell therapy, regeneration medicine

iHuman Institute: bio-imaging, biology, chemistry, computational biology, AI/ML

Institute of Mathematical Sciences: pure mathematics, theory of computing, applied mathematics

Institute of Humanities:Chinese philosophy, Western philosophy, logic, science philosophy, aesthetics, Ancient literature, modern literature, literary theory, comparative literature and world literature, Chinese writing, Chinese history, world history, historical theory, British and American language and literature, French language and literature, German language and literature, Japanese language and literature.

Following positions are opening:

1.Tenured and Tenure-track positions: assistant professor, associate professor and full professor.
Successful applicants will have a doctoral degree, and are expected to establish a record for independent, internationally recognized research, supervise students and teach high-quality courses.

2.Research positions: post-doctoral research fellow, research assistant professor, research associate professor and research professor.
Successful applicants will have a doctoral degree, a good research record and great passion for research.

3.Assistant positions: teaching assistant, research assistant, and administrative officer.
Successful applicants will have a Master's degree and relevant working experience.

ShanghaiTech will offer attractive compensation packages, including:

Initial research support package: reasonable start-up funds, research associates and post-doctoral fellows, laboratory space to meet research needs.

Compensation and benefits: highly competitive salary commensurate with experience and academic accomplishments, a comprehensive benefit package.

Subsidized housing: on-campus 80/100/120 m2 faculty apartments available at low rent for tenured and tenure-track faculty, on-campus postdoctoral dormitories, off-campus postdoctoral apartments and municipal apartments subsidized by Shanghai government.

Relocation & travel allowance: reimbursement of expenses for household relocation and family's one-way travel.

Family assistance: support with children's education; affiliated kindergarten, primary and middle schools.

To apply: using this format, please submit a cover letter (Firstname_Lastname_Cover_Letter.pdf), a research plan (Firstname_Lastname_Research_Plan.pdf), and a CV (Firstname_Lastname_CV.pdf) to shanghaitechuniversity@gmail.com

Biology of Henan University

Henan University, founded in 1912, is located in Kaifeng, a famous historic city which used to be the capital of China during eight different dynasties. In 2008, Henan University formally entered the list of the universities which are jointly developed by the provincial government and the ministry; in 2016, the University was selected in the "111 Plan"; and in 2017, the University becomes a "Double First-rate" university.

Henan University was formerly known as the Preparatory School for Further Study in Europe and America. It was later renamed Zhong zhou University, No.5 National Zhongshan University and Provincial Henan University. In 1942 its name was changed to National Henan University. After the founding of the People's Republic of China, its name has been changed many times. Finally, the name Henan University was restored in 1984.

For nearly 100 years since its founding, Henan University, adhering to its motto of "to illustrate illustrious virtue; to renovate the people; and to rest in the highest excellence", strengthening morality, has developed 600,000 kinds of talents. Among them, there are 57 academicians and members of the Social Science Academy.

Now Henan University is a comprehensive university with 12 branches of learning. It has established friendly relationship with more than 120 colleges in more than 40 countries and has become a member of the World Association of Universities and the Association of Universities of Asia and the Pacific. Henan University now has 33 Schools (Departments), 97 undergraduate programs, 42 primary discipline Master's programs, 20 professional Master's programs, 20 primary discipline doctoral program and has 15 doctoral research centers.

Initially founded in 1923, the School of Biological Sciences, and was re-established in 1987, and selected in the league of "Double First-rate" in 2017. The School of Biological Sciences delivers world-class research and teaching across a broad range of fundamental disciplines in biology. Specialized in "Plant Stress Biology and Sustainable Agriculture", leading researchers in School of Biological Sciences aims for exploring plant anti-stress potential, improving the utilization efficiency of water and nutrient, and enhancing pest control effectively etc., four major research interests in i) crop stress responses and signaling transduction, ii) crop development and stress adaptation, iii) biotic factor interactions in plant and ecological regulation, and iv) exploration of stress resistance gene and germplasm resources innovation. The school has been granted with more than 130 national research funding, including one National Program on Key Basic Research Project of China (973 Program), two National Major Scientific Research Program, two National Science Foundation for Distinguished Young Scholars, eight The General Program (Key Program) of National Natural Science Foundation of China, and over 120 National Natural Science Foundation of China grants. Research outputs were published in world-renowned peer-reviewed journals including Nature Communication, The Plant Cell, Plant Physiology etc.

The biology discipline is also the home base for State Key Laboratory for Cotton Biology, State Key Laboratory of Crop Stress Adaptation and Improvement, as well as Crop Stress Biology Overseas Expertise Introduction Project for Discipline Innovation (111 Plan). In addition, the school hosts Level I PhD Entitlement for Biological Sciences and Ecology and corresponding post-doctoral mobile station. The project "Stomatal Regulation Mechanism Enhancing Plant Water Efficiency" was awarded second prize in the National Nature Science Award 2012. The team initiated "Genetic Basis and Techniques for Biotic Water Saving" won Chang jiang Scholars and Innovative Research Team by Ministry of Education.

Except for focusing on basic scientific researches, the School of Biological Sciences developed inter-disciplinary researches, attempt to screen important anti-stress gene by using multi-omics, bioinformatics, genome modification and synthetic biology techniques, to breed new crop varieties featured with universal stress resistance, high-producing, high-quality through germplasm innovation and molecular design breeding.

We are eager for ambitious and outstanding researches, Welcome on board to be part of our history and our future.

Website: http://bio.henu.edu.cn
Email: wangg@henu.edu.cn

Hangzhou Innovation Research Institute of Beihang University

Hangzhou Innovation Research Institute of Beihang University is a new high-level research institute jointly established by Beihang University and the Zhejiang Province, Hangzhou City and Binjiang District governments. With the mission of "building a world-class technological innovation platform and innovative talent training platform in the field of information", and focusing on the multidisciplinary intersection of information technology, life and health, cognitive science and new materials, Hangzhou Innovation Research Institute actively explores new mechanisms and gathers global innovative resources, and is committed to achieving a number of major original innovations and key technological breakthroughs and applications, striving to become a talent and innovation center that is rooted into Zhejiang Province while looking to the world's first-class.

In March 2018, Hangzhou Innovation Research Institute of Beihang University officially settled in Binjiang District of Hangzhou, meaning the entering of the second 985 university in Hangzhou. The construction of its graduate school launched in September 2019 at Baima Lake of Binjiang District. The construction of hardware facilities is expected to be completed in about two years, and the scale of graduate students is expected to be 2,000. More importantly, Hangzhou Innovation Research Institute has undertaken the construction work of Sino-French Aviation University, providing teachers reserve for the university. On January 9, 2018, under the testimony of President of PRC Xi Jinping and French President Macron, Beihang University signed the memorandum of cooperation with Ecole Nationale de l'Aviation Civile (the French National Civil Aviation University) and agreed to jointly establish Sino-French Aviation University. The site of the university will be in the town of Pingyao, Yuhang District, Hangzhou, covering an area of 1,500 mu (10,000 acres).

I. "Qianjiang Forum"

The 3rd International Youth "Qianjiang Forum" of Hangzhou Innovation Research Institute of Beihang University will be held on December 28, 2019. The forum is aimed at exploring the hot topics of international academic frontiers in the information field through special reports, academic seminars and in-depth discussions.

II. Forum Setting

The forum sets up one main forum and four sub-forums naming Advanced Materials and Micro-processing Science, Instrument Science, Information Science and Technology, and Cyberspace Security Science. The paper awards session is also set up in the forum while each person is limited to submit one representative paper. The scope of the paper includes materials science and engineering, instrument science and technology, optical engineering, quantum precisive measurement, computer science and technology, software engineering, control science and engineering, information and communication engineering, cyberspace security and other relative fields. Representative papers should be published after 2015 as articles of the first author.

III. Application Conditions

(1) Those who have obtained a doctoral degree from a top-ranking university at home or abroad. After being selected, he/she can work full-time in the institute.

(2) Those who have innovative achievements acquired recognition at home and abroad. The achievements can be academic papers published in international leading journals or conferences in the relative fields, participation in international frontier research projects or national major scientific research tasks, and the main accomplishers of major scientific and technological awards.

IV. Application Method

Please visit the website http://hzii.buaa.edu.cn/info/1041/1664.htm and submit your application as required.

V. Contact Information

Contacts: Ms Tian, Mr Chen.
Contact Number: (+86)199-5789-0995
(+86)571-8536-7559
Email: buaahz_hr@buaa.edu.cn
Official Website: http://hzii.buaa.edu.cn

Information is power, and we are using our power to change the world.

Since established in 1927, the School of Information Management in Nanjing University has been recognized as one of the best information schools in China. As the third iSchools in Asia, SIM has been accredited as A+ level (highest level) and National Excellent Disciplines in China, and ranked 26th in QS World University Ranking 2019.

The School consists of 69 faculty and staff members, a number of well-known academic leaders. The School offers multi-level programs from undergraduate to PhD. By establishing four undergraduate, five graduate and doctoral specialties, we have currently built up a mature educational mode of cultivating compound talents within innovative consciousness and high practical ability.

Meanwhile, we attach great importance to scientific research as well as social service. Our faculty's innovative research is earning reputation and financial support from various foundations. During the last five years, we've completed over 160 academic programs and application projects collaborated with enterprises and organizations. We have also signed strategic cooperation agreements with many famous universities in the world and fully sharing the research resources of 65 international famous universities.

Now, the School of Information Management is searching for outstanding candidates to fill tenure/tenure-track faculty positions in Information Science, Library Science, Archive Management, Data Management or Digital Publishing research at the rank of Assistant, Associate, or Full Professor. Candidates must have a Ph.D. degree in information science, information systems, library science, archive science, computer science or a related discipline from accredited institutions. For tenure-track positions, candidates must also have a strong commitment to high-quality research and excellent teaching. For appointments at Assistant Professor level, candidates who are expected to complete their Ph.D. before enrollment are encouraged to apply. For appointments at the Associate or Full Professor level, a substantial record of research publications in top-tier journals is essential.

Preference will be given to candidates possessing a track record of publishing in leading journals in library and information science or information management. A record of obtaining grants from competitive institutions is also preferred. Successful candidates should have the ability to teach at both graduate and undergraduate levels. Shortlisted candidates will be contacted for interviews.

Applications including cover letter, curriculum vitae, referred publications, any available teaching evaluations, three letters of recommendation and availability for interviews, which should be submitted electronically to Ms. Ting Yu:

Ms. Ting Yu
Email: yuting@nju.edu.cn
Tel/Fax: 86-25-89687291
For more information,
see at: im.nju.edu.cn

Philosophy at Nanjing University

The Nanjing University department of philosophy ranks as one of China's foremost centers for philosophy research, it traces its roots back to its original establishment in 1920. In 2000, the department of religious studies was set up and it became one part of philosophy department. During the past years, a lot of famous scholars have been the faculty members of our department, such as Tang Yongtong, Zong Baihua, Fang Dongmei, etc.. In 1978, professor Hu Fuming, wrote an article "Practice is the only criterion of Truth", which made a very great impact to China's liberation of thought and the open-reform course.

Current available majors include: Marxist philosophy, chinese philosophy, foreign philosophy, ethics, logic, philosophy of science and technology, religious studies, eastern philosophy and religious studies. At present, the department has 56 teaching staff which contains 32 professors and 15 associate professors.

The department has produced a significant amount of published materials and academic literature appearing in the leading academic periodicals such as China Social Science and Philosophical Research. Since 1977 the philosophy department has trained more than 1400 undergraduates and several hundred graduate degree students. At present, the department has 135 undergraduates and almost 300 graduate students.

Website: https://www.nju.edu.cn/EN/80/51/c7136a163921/page.psp
Email: zhexue@nju.edu.cn

Law school of Nanjing University

Nanjing University is well known in China for its long history and high reputation. It was originally established in 1902. After its development for over a century, Nanjing University maintains good momentum among various fields and ranks top among higher educational institutions in China. It was designated as a Class A institution in the Double First Class University plan, a government initiative to cultivate an elite group of Chinese universities into "world-class" institutions. In 2018, it was ranked 114 on QS World University Rankings. Today, NJU has over 30000 full-time students in 31 schools and departments.

Law is one of the priority disciplines in NJU. Due to its commitment on facilitating legal education and research, Law School has cultivated many talents in law field, which has been well known nationally and internationally. There are 62 faculty members, including 27 professors, 29 associate professors and 6 researchers. It has 10 research institutions (e.g., the Sino-Deutsch Law Research Institute, the Chinese Law Cases Research Center, and Economic Law Research Institute, etc.) and a post-doctoral research station. Due to the efforts of its students and faculty, NJU Law School was evaluated as "A-" on the disciplines ranking of Chinese Ministry of Education.

For more information, please contact **swu@nju.edu.cn** (Siyue Wu, staff member of Personnel) or visit our website **law.nju.edu.cn**!

Recruitment of Global Talents for Guangdong Ocean University

Guangdong Ocean University (GDOU), located in an enchanting southernmost coastal city, Zhanjiang, in the mainland of China, is a key institution co-built by the People's Government of Guangdong Province and the State Oceanic Administration of China. Featuring Ocean and Fisheries Sciences and developing as a comprehensive, multidisciplinary one, GDOU has an integrated academic degree authorization system that provides bachelor's, master's, and doctoral degrees, and has been evaluated "excellent" by the Undergraduate Teaching Evaluation of the Ministry of Education of the P.R. China and crowned as a high-level university of key subject construction in Guangdong Province.

Our university has three campuses of different functions, 806 acres in total, of which the main campus is located at the east side of the National Class 4A Tourist Attraction, Huguangyan, one of the world renowned volcanic geo-parks. This dream campus, being a beautiful place full of trees and flowers, facing sea, and being surrounded by mountains, is home to studying, teaching, and researching.

GDOU prides herself on owning 3 first-class disciplines for doctoral degrees, 9 first-class master's degree programs, 44 secondary master's degree programs, 77 undergraduate majors, 1 comprehensive national reform pilot subject, 4 pilot subjects in the Educational Reform Plan of Excellent Agricultural Talents, 1 national off-campus practical education base for students, 33 research platforms at provincial or ministerial level, 13 provincial demonstration centers of experimental education, 21 key laboratories at department level, etc.

At GDOU, more than 35,000 students from different countries gather around to pursue their dreams. Over 2,500 qualified and brilliant staff and faculty are devoting themselves to teaching, researching, and more at GDOU.

Due to the fast development and establishment, GDOU is now recruiting high-level faculty from both China and abroad.

Disciplines

Sciences, Engineering, Economics, Management, Law, Literature, Education, marine-related disciplines, Art, etc. Please find the details in GDOU website.

Job Requirements

(1) PhD from both overseas and domestic universities/research institutes;

(2) Be capable of teaching and researching at university.

Contact us

Applicants could send a detailed CV to rcyj@gdou.edu.cn. Please detail your education, work experiences, publications, research interests, etc., and use Applicant's Name + Profession (Field of Research) + Current Institution of Study or Work + Position of interest as the subject of the email.

Tel: (+86) 0759-2383117
Email: rcyj@gdou.edu.cn

GDOU Website: www.gdou.edu.cn
Recruitment: http://zp.ehall.gdou.edu.cn/rsfw/sys/zp-glxt/extranet/index.do#/home

Southern Medical University Seeking Global Talents

Established in 1951 and located in the city of Guangzhou, the Southern Medical University (SMU) was designated as a national key university in 1979. SMU is the only university that ranks as a Ministries-Province Co-sponsored Medical University in Guangdong Province. As one of the eight universities, including PUMC, Peking University and Fudan University, approved by Ministry of Education to offer the eight-year clinical medicine program, SMU provides nationally recognized high-quality medical education and state-of-the-art biomedical and clinical research platforms and environment for faculties, students and trainees. At the post-graduate level, SMU admits over 2,300 doctoral and master students and hundreds of postdoc fellows each year.

With 275 National Natural Science Foundation of China (NSFC) grants awarded to SMU faculties in 2019, SMU has ranked among the top 27 ones of all Chinese universities receiving NSFC grants. Five disciplines of SMU are ranked top 1% in the world according to the Essential Science Indicators evaluation by the Web of Science. SMU has 8 first-levelpost-doctoral research stations, including Basic medicine, Clinical Medicine, Public Health and Preventive Medicine, The Integrative Medicine, Biology, Pharmaceutical Science, Biomedical Engineering and Nursing, and scientific fields with outstanding academic strength include molecular cell biology, stem cell and regenerative medicine, pathophysiological studies of human diseases, brain science and neurological diseases, nephrology, gastroenterology, hematology, infectious diseases, tropical diseases and bone/muscular diseases. Each year over 1200 research articles authored by SMU faculties are published in internationally circulating journals including JAMA, N Engl J Med, Nature Medicine, Nature Neuroscience, etc.

SMU owns a State Key Laboratory, a State Clinical Research Center and 75 province/ministries-designated research centers and laboratories. Eleven teaching hospitals are directly affiliated to SMU, receiving over 12000 patient visits and operate 200,000 surgeries annually. Together with the University and its schools/colleges, SMU hospitals are accommodating large numbers of researchers and are planning to continuously increase recruitment of new faculties at the ranks of full or associate professor in the next few years. Outstandingand motivated scholars are sought and welcome to join SMU in Guangzhou, an attractive city reputed for its being the frontier of Chinese reform and opening-up. ALL employed faculty at SMU receive an internationally competitive annual salary ranging from USD 60,000 to USD 179,000 , settling-in allowance ranging from USD 120,000 to USD 596,000 and comprehensive benefits package including adequate funding for research, necessary research facilities and office service supports, special quotas for graduate students and postdoc fellows, apartments and high quality healthcare service. Employed faculty will be appointed as a professor or associate professor, a postgraduate supervisor or a post-doctoral cooperative supervisor according to the level. SMU will also assist in research coordination, project applications and children's household registration. ALL post-doctor at SMU receive annual salary ranging from USD 42,500 to USD 85,000. Postdoctoral fellows can enjoy special allowance for published papers and funded projects during his/her stay at the Station, free Postdoctoral Apartment or housing allowance, the research subsidy provided by the cooperative advisors and the public resource services of the university. SMU will also offer opportunities for professional and technical title applications.

Application Documents

- CV with the basic information, education background and work experience, and main academic achievements.
- Attachments (optional for initial stages): Passport, doctorate degree certificate, overseas employment certificate, certificate for the main academic achievements.

Please send a copy of materials listed above to smurcb@126.com, please state clear information such as your names, disciplines or research fields in the title of your email.

Contact

For faculty positions

TEL: +86-20-61648600 Email: smurcb@126.com

Address: Talent Office, Southern Medical University, No.1023 Shatai Road, Guangzhou, Guangdong, 510515, PRC

For postdoctoral positions

Tel: +86-20-61648642 Email: smubgb@163.com

Address: Postdoctoral Management Office, Southern Medical University, No.1023 Shatai Road, Guangzhou, Guangdong, 510515, PRC

For more information please visit: http://www.smu.edu.cn/ or follow our wechat official account:

WORKING LIFE

By **Ashley Stenzel**

A mother's guilt

The sun was rising as we drove across the Minnesota state line, marking the moment my family and I left the only home we had ever known. I wanted to feel excited about my new Ph.D. program, but all I could feel was guilt. We were moving to New York so that I could pursue my goal of becoming a professor. The move was good for me professionally, but I worried about uprooting my husband and daughters. I also feared that—with the demands placed on me in grad school—I wouldn't be able to give my kids the childhood they deserved. The 3 years that have passed since then haven't been easy. But I've realized that I'm not the only person who benefits from my education. My kids do, too.

I was 17 years old when I learned that I was going to be a mother. Our first daughter came into the world 4 days after my high school graduation. I didn't know whether I was going to make it through college, let alone grad school. But my education was important to me because I'd witnessed my own mother attend college and advance her career when I was in high school. I wanted to follow in her footsteps.

I studied biology in college, taking a full course load and working night shifts at a local hospital to help provide for my new family. It was challenging to balance classes, work schedules, and being a mom. But I got through it, finding moments of joy along the way. On the nights when I was home, I'd read my class notes out loud with my daughter. She'd respond by asking questions, such as "Mom, what are bacteria?" It helped us both learn.

"Navigating academia as a young mother is hard, but it's also rewarding."

After I graduated, I knew that I would need a Ph.D. to land the kind of job I wanted. But I was nervous about whether the life of a grad student would be possible as a mother. I'd given birth to our second daughter during my last year of college, so we now had two young girls to raise.

When I interviewed for my Ph.D. position, I asked a senior grad student whether there were any resources to help student parents. "I'm sorry," she said. "I don't really know of anybody who would be able to help with that." Her answer reinforced a fear I'd harbored: that I would be a lone student parent in my Ph.D. cohort, trying to forge a path on my own.

I arrived in New York feeling more than the usual new-grad-student angst. I worried that I wouldn't be able to cut it academically, and it didn't help that—as I'd feared—none of my peers were parents. I also suffered from a more personal fear that I was being selfish—that my decision to prioritize my career was going to have long-term negative repercussions for my kids. I imagined them looking back in 20 years and thinking that I didn't spend enough time with them, or that I took away their happiness.

So, I made a rule to never be visibly upset about my work in front of my children. When I went home, they needed me to just be their mom. Grad school was stressful, but it felt unfair to complain about a life that I had asked my family to sacrifice for. One night after a tough day at work, I pulled into our driveway, sat in the car, and let a few tears fall down my cheeks. Then, I pulled myself together and put on a smile when I greeted my daughters at our front door.

Over the past year, though, I've started to let go of some of this worry. I've realized that we have not only adapted to our new situation, but we are thriving. My husband landed a job that he is happy with. My older daughter dreams of becoming a marine biologist. And my younger daughter loves exploring, something we do often now that we live in a new state. Both daughters also constantly remind me that they're proud of the things I do. Recently, while driving past the cancer institute I work at, my older daughter said, "Thinking about people having cancer is so sad, but I feel better knowing that you are researching it to help."

Navigating academia as a young mother is hard, but it's also rewarding. My kids are learning to look at the world through the lens of science, and watching their mom succeed inspires them. I look forward to seeing them follow my footsteps, whatever path they choose. ■

Ashley Stenzel is a Ph.D. student at the University at Buffalo, part of the State University of New York system.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS